

# LP Bounds for Rate-Distortion with Variable Side Information

Sinem Unal and Aaron B. Wagner

## Abstract

We consider a rate-distortion problem with side information at multiple decoders. Several upper and lower bounds have been proposed for this general problem or special cases of it. We provide an upper bound for general instances of this problem, which takes the form of a linear program, by utilizing random binning and simultaneous decoding techniques [1] and compare it with the existing bounds. We also provide a lower bound for the general problem, which was inspired by a linear-programming lower bound for index coding, and show that it subsumes most of the lower bounds in literature. Using these upper and lower bounds, we explicitly characterize the rate-distortion function of a problem that can be seen as a Gaussian analogue of the “odd-cycle” index coding problem.

## I. INTRODUCTION

We consider the rate-distortion tradeoff for a canonical problem in source coding: an encoder with access to a source of interest broadcasts a single message to multiple decoders, each endowed with side information about the source. Each decoder then wants to reproduce the source subject to a distortion constraint. This is essentially the multiple-decoder extension of the Wyner-Ziv [2] problem, sometimes referred to as the *Heegard-Berger* [3] problem.

Even for the two-decoder case, the complete characterization of the rate-distortion function is a long-standing open problem. However, the rate-distortion function has been determined in several special cases, including when the side information at the various decoders can be ordered according to stochastic degradedness [3], when there are two decoders whose side information is “mismatch degraded” [4], and when there are two decoders and the side information at decoder 2 is “conditionally less noisy” than the side information at decoder 1 and decoder 1 seeks to losslessly reproduce a deterministic function of the source [5]. Also, instead of imposing some degraded structure on the side information, one can consider degraded reconstruction sets at the two decoders in which one component of the source is reconstructed at both decoders with vanishing block error probability and the other component of the source is only reconstructed at a single decoder [6]. Various vector Gaussian instances of the problem are solved [7], [8]. Several instances of the index coding problem, which is an important special case, have also been solved (e.g., [9]–[11]).

Upper and lower bounds on the rate-distortion function in the general case are also available. Existing achievable schemes proceed by crafting separate messages for different subsets of decoders, which are encoded and decoded in a fixed order using random binning [3], [4], [12]. Our first contribution is to show how such schemes can be improved using *simultaneous decoding* [1], in which each decoder decodes all of its messages at once instead of sequentially. The resulting achievable bound involves optimizing over auxiliary random variables and, for each choice of such variables, solving a linear program (LP). Prior to this work, the best achievable bound was due to Timo *et al.* [12], who corrected an earlier achievable bound due to Heegard and Berger [3]. In fact, as we discuss in Section V, the proof given by Timo *et al.* contains an error similar to the one contained in Heegard and Berger.

One natural way of obtaining a lower bound is to consider a relaxed instance of the problem in which the side information at some of the decoders is enhanced in such a way that the problem becomes stochastically degraded. Indeed, most existing lower bounds adopt this approach in some form [8], [11]. For the special case of index coding, Blasiak *et al.* [10] provide a lower bound that takes the form of a linear program, the constraints for which are derived from properties of the entropy functional, such as submodularity. This raises the question of whether a similar-style bound can be obtained for more general instances of the problem. The second main contribution of the paper is such a bound. It is obtained by introducing a notion of *generalized side information* and capturing the properties of mutual information in the form of a linear program. We show that this lower bound subsumes several existing lower bounds.

To demonstrate the efficacy of our upper and lower bounds, we consider a rate-distortion problem obtained by extending the odd-cycle index coding problem to Gaussian sources with mean squared error (MSE) distortion constraints. We find an explicit expression for its rate-distortion function by combining the two bounds.

The outline of the paper is as follows. Section II formulates the general rate-distortion problem. Section III presents the LP-type upper bound based on simultaneous decoding while Section III-A provides the extension of this upper bound to

S. Unal was with Cornell University, School of Electrical & Computer Engineering, Ithaca, NY 14853 USA. She is now with KenCast Inc, Norwalk, CT 06854 USA. A. B. Wagner is with Cornell University, School of Electrical & Computer Engineering, Ithaca, NY 14853 USA (e-mail: su62@cornell.edu, wagner@ece.cornell.edu). This paper was presented in part at the IEEE Int. Symposium on Information Theory (ISIT), Barcelona, July 2016 and it is submitted for presentation in part to Data Compression Conference (DCC), Snowbird, UT April, 2017.

Gaussian sources. In Section IV, we provide the LP-type lower bound and in Section V we show that the LP-type upper and lower bounds subsume several existing bounds. Lastly Section VI presents optimality results including the rate distortion characterization of the odd-cycle Gaussian problem.

## II. PROBLEM DESCRIPTION

Let  $X$  denote the source at the encoder and  $\mathcal{X}$  denote the source alphabet. Also,  $Y_l \in \mathcal{Y}_l$ ,  $l \in [m]$  denotes the side information at decoder  $l$  and  $Y_l$  is jointly distributed with the source,  $X$ . Lastly,  $\hat{X}_l \in \hat{\mathcal{X}}_l$  denotes the reconstruction of  $X$  at decoder  $l$  and  $D_l$  denotes the corresponding distortion constraint. Each decoder wishes to reconstruct the source,  $X$ , subject to its distortion constraint and we assume initially that the source alphabet,  $\mathcal{X}$ , the side information alphabets,  $\mathcal{Y}_l$ ,  $l \in [m]$ , and the reconstruction alphabets  $\hat{\mathcal{X}}_l$ ,  $l \in [m]$ , are finite. We use the following definitions to formulate the problem.

**Definition 1.** An  $(n, M, \mathbf{D})$  code where  $n$  denotes the blocklength and  $M$  denotes the message size and  $\mathbf{D} = (D_1, \dots, D_m)$  is composed of

- an encoding function

$$f : \mathcal{X}^n \rightarrow \{1, \dots, M\}$$

- and decoding functions

$$\begin{aligned} g_1 &: \{1, \dots, M\} \times \mathcal{Y}_1^n \rightarrow \hat{\mathcal{X}}_1^n \\ &\vdots \\ g_m &: \{1, \dots, M\} \times \mathcal{Y}_m^n \rightarrow \hat{\mathcal{X}}_m^n \end{aligned}$$

satisfying the distortion constraints

$$E \left[ \frac{1}{n} \sum_{k=1}^n d_l(X_k, \hat{X}_{lk}) \right] \leq D_l, \text{ for } l \in [m]$$

where

$$\hat{X}_l^n = g_l(f(X^n), Y_l^n), \text{ for } l \in [m]$$

and  $d_l(\cdot, \cdot) \in [0, \infty)$  is the distortion measure for decoder  $l$ .

**Definition 2.** A rate  $R$  is  $\mathbf{D}$ -achievable if for every  $\epsilon > 0$  there exists an  $(n, M, \mathbf{D} + \epsilon \mathbf{1})$  (where  $\mathbf{1}$  is the all-ones vector) code such that for sufficiently large  $n$  we have  $n^{-1} \log M \leq R + \epsilon$ .

We define the rate-distortion function as

$$R(\mathbf{D}) = \inf \{ R : R \text{ is } \mathbf{D}\text{-achievable} \}.$$

## III. SIMULTANEOUS DECODING BASED UPPER BOUND TO $R(\mathbf{D})$

We present our first main result, which is an upper bound to the rate-distortion function  $R(\mathbf{D})$ . The following notation, which is similar to that in [12], will be useful to state the results.

**Notation 1.** Let  $(X, Y, Z)$  be a random vector. Then  $X \perp Y$  denotes that  $X$  and  $Y$  are independent,  $X \perp Y | Z$  denotes that  $X$  and  $Y$  are conditionally independent given  $Z$ , and  $X \leftrightarrow Y \leftrightarrow Z$  denotes that  $X$ ,  $Y$  and  $Z$  form a Markov chain.

**Notation 2.**  $v = \mathcal{S}_1, \dots, \mathcal{S}_{2^m-1}$  denotes an ordered list of all nonempty subsets of  $[m]$ , where each  $\mathcal{S}_i$  denotes a different subset.  $\mathcal{V}$  denotes the set of all possible such  $v$ .

**Notation 3.** Let  $v \in \mathcal{V}$  be fixed.  $\mathcal{U}_{\mathcal{S}_1}, \dots, \mathcal{U}_{\mathcal{S}_{2^m-1}}$  denote the alphabets of finite-alphabet random variables  $U_{\mathcal{S}_1}, \dots, U_{\mathcal{S}_{2^m-1}}$  respectively.  $\mathcal{P}_v$  denotes the set of all distributions on  $\mathcal{U}_v^* \times \mathcal{X} \times \mathcal{Y}^*$  where  $\mathcal{U}_v^* = \mathcal{U}_{\mathcal{S}_1} \times \dots \times \mathcal{U}_{\mathcal{S}_{2^m-1}}$  and  $\mathcal{Y}^* = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$ .

**Notation 4.** Let  $\mathcal{U} = \{U_{\mathcal{S}_1}, U_{\mathcal{S}_2}, \dots, U_{\mathcal{S}_{2^m-1}}\}$ ,  $\mathcal{D}_l = \{\mathcal{S}_i | l \in \mathcal{S}_i\}$ , and  $\mathcal{D}'_l$  be a nonempty subset of  $\mathcal{D}_l$ . Then we define

$$\begin{aligned} U_{\mathcal{D}'_l} &= \{U_{\mathcal{S}_i} \in \mathcal{U} | \mathcal{S}_i \in \mathcal{D}'_l\} \\ U_{\mathcal{S}_j}^- &= \{U_{\mathcal{S}_i} \in \mathcal{U} | i < j\} \\ U_{\mathcal{S}_j, \mathcal{D}'_l}^- &= \{U_{\mathcal{S}_i} \in U_{\mathcal{S}_j}^- | \mathcal{S}_i \in \mathcal{D}'_l\}. \end{aligned}$$

**Theorem 1.** The rate-distortion function,  $R(\mathbf{D})$ , is upper bounded by

$$R_{ach}(\mathbf{D}) = \text{conv}(R'_{ach}(\mathbf{D})) \quad (1)$$

where  $\text{conv}(R'_{ach}(\mathbf{D}))$  denotes the lower convex envelope of  $R'_{ach}(\mathbf{D})$  with respect to  $\mathbf{D}$ ,

$$R'_{ach}(\mathbf{D}) = \min_{v \in \mathcal{V}} \inf_{C_{ach,v}(\mathbf{D})} \inf_{C_{ach}^{LP}} \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j}, \quad (2)$$

$C_{ach,v}(\mathbf{D})$  denotes the set of  $p \in \mathcal{P}_v$  such that

- 1)  $p(x, y_1, \dots, y_m)$  equals the joint distribution of  $(X, Y_1, \dots, Y_m)$
- 2)  $\mathcal{U} \leftrightarrow X \leftrightarrow (Y_1, \dots, Y_m)$
- 3) There exist functions  $g_l(U_{\mathcal{D}_l}, Y_l)$  such that  $E[d_l(X, g_l(U_{\mathcal{D}_l}, Y_l))] \leq D_l$  for all  $l \in [m]$ ,

and

$C_{ach}^{LP}$  denotes the set of  $R_{\mathcal{S}_j}, R'_{\mathcal{S}_j}$ , where  $\mathcal{S}_j \in v$ , such that

- 1)  $R_{\mathcal{S}_j} \geq 0, R'_{\mathcal{S}_j} \geq 0$  for all  $j \in [2^m - 1]$
- 2)  $R_{\mathcal{S}_j} \geq I(X, U_{\mathcal{S}_j}^-; U_{\mathcal{S}_j}) - R'_{\mathcal{S}_j}$  for all  $j \in [2^m - 1]$
- 3) For each decoder  $l, l \in [m]$ ,  

$$\sum_{\mathcal{S}_j \in \mathcal{D}_l'} R'_{\mathcal{S}_j} \leq \sum_{\mathcal{S}_j \in \mathcal{D}_l'} H(U_{\mathcal{S}_j}) - H(U_{\mathcal{D}_l'} | U_{\mathcal{D}_l \setminus \mathcal{D}_l'}, Y_l), \text{ for all } \mathcal{D}_l' \subseteq \mathcal{D}_l.$$

*Proof of Theorem 1:* The proof is given in the Appendix A. ■

**Remark 1.** Using the chain rule, we can rewrite condition 3) of  $C_{ach}^{LP}$  in Theorem 1 as

$$\begin{aligned} & \text{for each decoder } l, l \in [m], \\ & \sum_{\mathcal{S}_j \in \mathcal{D}_l'} R'_{\mathcal{S}_j} \leq \sum_{\mathcal{S}_j \in \mathcal{D}_l'} I(U_{\mathcal{S}_j}; U_{\mathcal{S}_j, \mathcal{D}_l'}^-, U_{\mathcal{D}_l \setminus \mathcal{D}_l'}, Y_l), \text{ for all } \mathcal{D}_l' \subseteq \mathcal{D}_l. \end{aligned} \quad (3)$$

This representation will be useful when we extend this theorem to continuous sources. Hence, from now on we consider the condition 3) of  $C_{ach}^{LP}$  in the form of (3).

**Remark 2.** Since  $R_{ach}(\mathbf{D})$  is an upper bound to the rate-distortion function,  $R(\mathbf{D})$ , we can obtain a computable upper bound to  $R(\mathbf{D})$  by imposing a cardinality constraint on the alphabets of auxiliary random variables  $U_{\mathcal{S}_j}$  in Theorem 1.

The scheme can be described as follows. Each  $U_{\mathcal{S}_j}$  in Theorem 1 can be viewed as a message for all decoders  $l$  for which  $l \in \mathcal{S}_j$ . The encoder encodes each message  $U_{\mathcal{S}_j}$  with respect to the order  $v \in \mathcal{V}$ , using random binning. Here  $R_{\mathcal{S}_j}$  and  $R'_{\mathcal{S}_j}$  can be interpreted as the number of bins in the codebook of message  $U_{\mathcal{S}_j}$  and the number of codewords per bin, respectively. Then each decoder  $l$  decodes its messages using simultaneous decoding and reconstructs the source using these messages and its side information,  $Y_l$ .

#### A. Rate-Distortion Function with Gaussian Source and Side Information

We next extend the achievable scheme in Theorem 1 to the rate-distortion problem with vector Gaussian sources. More specifically, we are interested in the following rate-distortion problem. The source and the side information at the decoders,  $(\mathbf{X}, \mathbf{Y}_1, \dots, \mathbf{Y}_m)$ , are zero mean jointly Gaussian vectors. The source  $\mathbf{X} = (X_1, \dots, X_k)$  has length  $k$  and the length of  $\mathbf{Y}_i$  is  $k_i, i \in [m]$ .

**Notation 5.** Let  $\mathbf{v}$  and  $\mathbf{w}$  be  $k \times 1$  vectors. The notation  $\mathbf{v} \leq \mathbf{w}$  denotes that the  $i^{\text{th}}$  component of  $\mathbf{v}$ , denoted by  $v_i$ , is less than or equal to  $w_i$  for all  $i \in [k]$ .

**Notation 6.** Let  $M$  be an  $m \times m$  matrix.  $(M)_d$  denotes the vector whose  $i^{\text{th}}$  component is the  $i^{\text{th}}$  diagonal element of  $M$ ,  $i \in [m]$ .

**Notation 7.**  $K_{\mathbf{X}}$  denotes the covariance matrix of  $\mathbf{X}$ .  $K_{\mathbf{X}|\mathbf{Y}}$  denotes the conditional covariance matrix of  $\mathbf{X}$  conditioned on  $\mathbf{Y}$ .

Let  $\mathbf{D}_i > 0$  for all  $i \in [m]$ . The distortion constraints are

$$\left( \frac{1}{n} \sum_{k=1}^n E \left[ (\mathbf{X}_k - \hat{\mathbf{X}}_{\mathbf{i}k})(\mathbf{X}_k - \hat{\mathbf{X}}_{\mathbf{i}k})^T \right] \right)_d \leq \mathbf{D}_i, \text{ for all } i \in [m], \quad (4)$$

i.e., we impose component-wise mean square error ( $MSE$ ) distortion constraints. Since we have  $MSE$  distortion constraints, without loss of generality we can take the reconstruction at each decoder to be the conditional expectation of the source given the output of the encoder and the corresponding side information. From now on, we denote the rate-distortion function of this problem as  $R^{MSE}(\mathbf{D})$ .

**Theorem 2.** *Let the joint distribution of  $(\mathbf{X}, \mathbf{Y}_i)$ ,  $i \in [m]$  be given. Then the rate-distortion function,  $R^{MSE}(\mathbf{D})$ , is upper bounded by*

$$R_{ach}^G(\mathbf{D}) = \min_{v \in \mathcal{V}} R_{ach,v}^G(\mathbf{D}) \quad (5)$$

where

$$R_{ach,v}^G(\mathbf{D}) = \inf_{C_{ach,v}^G(\mathbf{D})} \inf_{C_{ach}^{LP}} \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j}$$

$C_{ach,v}^G(\mathbf{D})$  denotes the set of  $p \in \mathcal{P}_v$  such that

- 1)  $p(\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_m)$  equals the joint distribution of  $(\mathbf{X}, \mathbf{Y}_1, \dots, \mathbf{Y}_m)$
- 2)  $T$  is a discrete random variable over  $[\tau]$  for some positive integer  $\tau$  such that  $T \perp (\mathbf{X}, \mathbf{Y}_1, \dots, \mathbf{Y}_m)$
- 3)  $\mathcal{U} \leftrightarrow \mathbf{X} \leftrightarrow (\mathbf{Y}_1, \dots, \mathbf{Y}_m)$
- 4)  $K_{\mathbf{X}|\mathbf{U}_{\mathcal{S}_i}, \mathbf{Y}_i} \leq \mathbf{D}_i$  for all  $i \in [m]$
- 5)  $\mathbf{U}_{\mathcal{S}_j} = (\mathbf{U}_{\mathcal{S}_j,t}, T)$  such that  $\mathbf{U}_{\mathcal{S}_j} = \mathbf{U}_{\mathcal{S}_j,t}$  if  $T = t$ , all  $\mathbf{U}_{\mathcal{S}_j,t}$  are jointly Gaussian with  $(\mathbf{X}, \mathbf{Y}_1, \dots, \mathbf{Y}_m)$ , and  $I(\mathbf{U}_{\mathcal{S}_j}; \mathbf{U}_{\mathcal{S}_i}, \mathbf{X}) < \infty$  for all  $\mathcal{S}_j \in [2^m - 1]$ ,  $\mathcal{S}_i \in [2^m - 1]$  and  $i \neq j$ ,

and  $C_{ach}^{LP}$  is the set of conditions obtained by replacing each  $X$ ,  $Y_i$ , and  $U_{\mathcal{S}_j}$  in the conditions of  $C_{ach}^{LP}$  in Theorem 1 by  $\mathbf{X}$ ,  $\mathbf{Y}_i$ , and  $\mathbf{U}_{\mathcal{S}_j}$  respectively.

**Remark 3.** Since all feasible messages  $\mathbf{U}_{\mathcal{S}_j}$  in (5) are Gaussian mixtures and the source and the side information vectors are jointly Gaussian, the minimum mean square error (MMSE) estimator is linear given the realization of  $T$ . In other words, we can write  $\hat{\mathbf{X}}_i = A_{i,t} \mathbf{U}_{\mathcal{S}_i} + B_{i,t} \mathbf{Y}_i$  given  $T = t$ , where the value of  $A_{i,t}$  and  $B_{i,t}$  are determined by the joint distribution  $p \in C_{ach,v}^G(\mathbf{D})$ .

*Proof of Theorem 2:* The argument is based on a quantization of the source and messages similar to the procedure in [1, Section 3]. First we quantize the source, all messages and the side information. Then we apply the achievable scheme in the proof of Theorem 1 to these quantized variables and show that the rate in (5) is  $\mathbf{D}$ -achievable for our problem.

Let  $v \in \mathcal{V}$  be fixed and  $\epsilon > 0$  be given. Also let  $(\mathbf{X}, \mathcal{U}, \mathbf{Y}_1, \dots, \mathbf{Y}_m)$  be such that the joint distribution of it, denoted by  $p$ , is in  $C_{ach,v}^G(\mathbf{D})$ . Note that we can represent each message  $\mathbf{U}_{\mathcal{S}_j} = A_{\mathcal{S}_j,t} \mathbf{X} + \mathbf{N}_{\mathcal{S}_j,t}$ ,  $\mathcal{S}_j \in v$ , if  $T = t$  where  $\mathbf{N}_{\mathcal{S}_j,t} \perp (\mathbf{Y}_1, \dots, \mathbf{Y}_m)$  and we can represent the side information as  $\mathbf{Y}_i = B_i \mathbf{X} + \mathbf{N}_i$ ,  $i \in [m]$  where  $\mathbf{N}_i \perp (\mathcal{U}, \mathbf{X})$ . Now we quantize  $\mathbf{X}$  and all  $\mathbf{Y}_i$ ,  $i \in [m]$ , and we use the notation  $\bar{\mathbf{X}}$  to denote the quantized version of  $\mathbf{X}$ . We perform the quantization such that

$$E[(X_j - \bar{X}_j)^2] \leq \delta(\epsilon) \min_{i \in [m]} D_{ij} \text{ for all } j \in [k] \quad (6)$$

$$E[(\bar{X}_j - \widehat{\bar{X}}_{ij})^2] \leq D_{ij} + \delta(\epsilon) D_{ij} \text{ for all } i \in [m] \text{ and } j \in [k] \quad (7)$$

$$|I(\mathbf{X}, \mathbf{U}_{\mathcal{S}_j}^-; \mathbf{U}_{\mathcal{S}_j}) - I(\bar{\mathbf{X}}, \bar{\mathbf{U}}_{\mathcal{S}_j}^-; \bar{\mathbf{U}}_{\mathcal{S}_j})| \leq \delta(\epsilon), \text{ for all } \mathcal{S}_j \in [2^m - 1] \quad (8)$$

$$\left| \sum_{\mathcal{S}_j \in \mathcal{D}'_i} I(\mathbf{U}_{\mathcal{S}_j}; \mathbf{U}_{\mathcal{S}_j, \mathcal{D}'_i}^-, \mathbf{U}_{\mathcal{D}_i \setminus \mathcal{D}'_i}, \mathbf{Y}_i) - \sum_{\mathcal{S}_j \in \mathcal{D}'_i} I(\bar{\mathbf{U}}_{\mathcal{S}_j}; \bar{\mathbf{U}}_{\mathcal{S}_j, \mathcal{D}'_i}^-, \bar{\mathbf{U}}_{\mathcal{D}_i \setminus \mathcal{D}'_i}, \bar{\mathbf{Y}}_i) \right| \leq \delta(\epsilon), \text{ for all } i \in [m] \text{ and } \mathcal{D}'_i \subseteq \mathcal{D}_i, \quad (9)$$

where  $\delta(\epsilon) > 0$  is to be specified later, and

$$\bar{\mathcal{U}} \leftrightarrow \bar{\mathbf{X}} \leftrightarrow \mathbf{X} \leftrightarrow (\mathbf{Y}_1, \dots, \mathbf{Y}_m) \leftrightarrow (\bar{\mathbf{Y}}_1, \dots, \bar{\mathbf{Y}}_m).$$

Let  $\bar{p}$  denote the joint distribution of  $(\bar{\mathcal{U}}, \bar{\mathbf{X}}, \bar{\mathbf{Y}}_1, \dots, \bar{\mathbf{Y}}_m)$ . Now we form a new problem in which the source is  $\bar{\mathbf{X}}$ , the side information at decoder  $i$  is  $\bar{\mathbf{Y}}_i$ ,  $i \in [m]$ , and the distortion constraints are as in (7). Note that for this problem,  $\bar{p}$  is in  $C_{ach,v}((1 + \delta(\epsilon))\mathbf{D})$  in (2). Then we can apply the achievable scheme in the proof Theorem 1 to the new problem.

Let  $R_{LP}((1+\delta(\epsilon))\mathbf{D}, \bar{p})$  denote the result of the linear program  $\inf_{C_{ach}^{LP}} \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j}$  in Theorem 1 when the joint distribution is  $\bar{p}$ . Then from Theorem 1, rate  $R_{LP}((1+\delta(\epsilon))\mathbf{D}, \bar{p})$  is  $(1+\delta(\epsilon))\mathbf{D}$ -achievable for the new problem. In other words, we can find an  $(n, M, (1+\delta(\epsilon))\mathbf{D} + \epsilon'\mathbf{1})$ ,  $\epsilon'(\epsilon) > 0$  (specified later), code with rate

$$R_{LP}((1+\delta(\epsilon))\mathbf{D}, \bar{p}) + \epsilon'(\epsilon) \quad (10)$$

and

$$\left( \frac{1}{n} \sum_{j=1}^n E \left[ (\bar{\mathbf{X}}_j - \widehat{\bar{\mathbf{X}}}_{i_j})(\bar{\mathbf{X}}_j - \widehat{\bar{\mathbf{X}}}_{i_j})^T \right] \right)_d \leq (1+\delta(\epsilon))\mathbf{D}_i + \epsilon'\mathbf{1} \text{ for all } i \in [m] \quad (11)$$

when the blocklength,  $n$ , is sufficiently large.

For our original problem, first we quantize the source, the side information and all the messages distributed by  $p$  as described above and then we apply the  $(n, M, (1+\delta(\epsilon))\mathbf{D} + \epsilon'\mathbf{1})$  code with rate (10) to these quantized variables, the joint distribution of which is  $\bar{p}$ . Let  $R_{LP}^G(\mathbf{D}, p)$  denote the result of the linear program  $\inf_{C_{ach}^{LP}} \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j}$  in Theorem 2 when the joint distribution is  $p$ . Note that the linear programs defining both  $R_{LP}((1+\delta(\epsilon))\mathbf{D}, \bar{p})$  and  $R_{LP}^G(\mathbf{D}, p)$  are finite. Thus by (8), (9) and standard results on the continuity of linear programs [13], we have that

$$|R_{LP}^G(\mathbf{D}, p) - R_{LP}((1+\delta(\epsilon))\mathbf{D}, \bar{p})| \leq \gamma(\epsilon),$$

where  $\gamma(\epsilon) \rightarrow 0$  as  $\delta(\epsilon) \rightarrow 0$ . Lastly utilizing the Cauchy and Jensen inequalities and using (6) and (11) as in [1, Section 3], we can obtain<sup>1</sup>

$$\begin{aligned} & \left( \frac{1}{n} \sum_{j=1}^n E \left[ (\mathbf{X}_j - \widehat{\mathbf{X}}_{i_j})(\mathbf{X}_j - \widehat{\mathbf{X}}_{i_j})^T \right] \right)_d \\ & \leq \delta(\epsilon)\mathbf{D}_i + (1+\delta(\epsilon))\mathbf{D}_i + \epsilon'\mathbf{1} + 2\sqrt{(\delta(\epsilon)\mathbf{D}_i((1+\delta(\epsilon))\mathbf{D}_i + \epsilon'\mathbf{1}))} \\ & = \mathbf{D}_i + 2\delta(\epsilon)\mathbf{D}_i + \epsilon'\mathbf{1} + 2\sqrt{(\delta(\epsilon)\mathbf{D}_i((1+\delta(\epsilon))\mathbf{D}_i + \epsilon'\mathbf{1}))} \text{ for all } i \in [m], \end{aligned} \quad (12)$$

for sufficiently large  $n$ .

Thus for all sufficiently large  $n$ , there exists a code whose rate does not exceed

$$R_{LP}^G(\mathbf{D}, p) + \epsilon'(\epsilon) + \gamma(\epsilon)$$

and whose distortion at the decoder  $i$  is dominated by the expression in (12). It follows that  $R_{LP}^G(\mathbf{D}, p)$  is  $\mathbf{D}$ -achievable. ■

#### IV. AN LP LOWER BOUND TO $R(\mathbf{D})$

We present our second main result, a lower bound on the rate-distortion function  $R(\mathbf{D})$  of the problem where the source  $\mathbf{X}$  and side information  $\mathbf{Y}_i$  are random vectors and the distortion constraint for each decoder  $i$  is  $\mathbf{d}_i(\mathbf{X}, \widehat{\mathbf{X}}_i) \leq \mathbf{D}_i$ . The same definitions for the scalar case are used to formulate this problem by replacing the scalar source, side information and distortion constraints by the vector ones given above.

**Definition 3.** [5]  $\mathbf{B}$  is conditionally less noisy than  $\mathbf{A}$  given  $\mathbf{C}$ , denoted as  $(\mathbf{B} \succeq \mathbf{A}|\mathbf{C})$ , if  $I(\mathbf{W}; \mathbf{B}|\mathbf{C}) \geq I(\mathbf{W}; \mathbf{A}|\mathbf{C})$  for all  $\mathbf{W}$  such that  $\mathbf{W} \leftrightarrow (\mathbf{X}, \mathbf{C}) \leftrightarrow (\mathbf{A}, \mathbf{B})$ .

**Definition 4.** Given a random vector  $\mathbf{W}$ ,  $\mathcal{C}(\mathbf{W})$  denotes the set of joint distributions over two vectors where the first vector has the same marginal distribution as  $\mathbf{W}$ .

We informally refer to  $\mathcal{C}(\mathbf{W})$  as the “set of random vectors coupled to  $\mathbf{W}$ ” and we sometimes write  $\mathbf{V} \in \mathcal{C}(\mathbf{W})$  to denote such a random vector.

**Definition 5.** Given  $\mathbf{V} \in \mathcal{C}(\mathbf{X})$  and a mapping  $\mathbf{U} : \mathcal{C}(\mathbf{X}) \rightarrow \mathcal{C}(\mathbf{X}, \mathbf{V})$ , let  $R_{lb}^{LP}(\epsilon)$  denote the infinite-dimensional LP in Table I, where  $K(\cdot)$  varies over all maps from  $\mathcal{C}(\mathbf{X})$  to  $[0, \infty)$ , and  $f_1(\cdot)$  and  $f_2(\cdot)$  are deterministic functions. Here  $K(\cdot)$  assigns the same number to all deterministic random variables and  $K(\emptyset)$  denotes this common number. Whenever  $(\mathbf{U}_A, \mathbf{V}, \mathbf{X}, \mathbf{A}, \mathbf{B})$  appear together, their joint distribution is assumed to satisfy  $(\mathbf{U}_A, \mathbf{V}) \leftrightarrow \mathbf{X} \leftrightarrow (\mathbf{A}, \mathbf{B})$ .

**Theorem 3.** For any  $\epsilon > 0$ ,  $R(\mathbf{D})$  is lower bounded by

$$R_{lb}(\mathbf{D} + \epsilon\mathbf{1}) = \inf_{\mathbf{V} \in \mathcal{C}(\mathbf{X})} \inf_{\mathbf{U} : \mathcal{C}(\mathbf{X}) \rightarrow \mathcal{C}(\mathbf{X}, \mathbf{V})} R_{lb}^{LP}(\epsilon) \quad (13)$$

<sup>1</sup>When  $\mathbf{v}, \mathbf{w}$  are  $k \times 1$  vectors,  $\mathbf{u} = \mathbf{v}\mathbf{w}$  is also a  $k \times 1$  vector such that  $u_i = v_i w_i$ ,  $i \in [k]$ .

TABLE I  
LP FOR THE RATE-DISTORTION PROBLEM

$\inf K(\emptyset) - \epsilon$  subject to  
 $K(\mathbf{X}) = 0$  (initialize)  
 $K(\mathbf{A}) \geq 0$ , for all  $\mathbf{A}$  (non-negativity)  
 $K(\mathbf{B}) + I(\mathbf{B}; \mathbf{V}, \mathbf{U}_{\mathbf{B}}|\mathbf{A}) \geq K(\mathbf{A})$ , for all  $(\mathbf{A}, \mathbf{B})$  couplings :  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$  (slope)  
 $K((\mathbf{A}, \mathbf{C})) \geq K((\mathbf{B}, \mathbf{C}))$ , for all  $(\mathbf{A}, \mathbf{B})$  couplings :  $(\mathbf{B} \succeq \mathbf{A}|\mathbf{C})$  (monotonicity)  
 $K(\mathbf{A}) \geq K(\mathbf{B}) + I(\mathbf{B}; \mathbf{V}, \mathbf{U}_{\mathbf{A}}|\mathbf{A})$ , for all  $(\mathbf{A}, \mathbf{B})$  couplings :  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$  (monotonicity+)  
 $K(\mathbf{A}) + K(\mathbf{B}) \geq K(\mathbf{C}) + K((\mathbf{A}, \mathbf{B}))$ , for all  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$  couplings :  $\mathbf{B} \leftrightarrow \mathbf{C} \leftrightarrow \mathbf{A}$  and  $\mathbf{C} = f_1(\mathbf{A})$  or  $\mathbf{C} = f_2(\mathbf{B})$  (submodularity)

where  $\mathbf{V}$  and  $\mathbf{U}$ . in the infima must satisfy

- 1) For all  $\mathbf{B} \in \mathcal{C}(\mathbf{X})$ ,  $\mathbf{U}_{\mathbf{B}}$  is independent of  $\mathbf{X}$ .
- 2) If  $\{\mathbf{B}, A_1, \dots, A_s\}$  are all elements of  $\mathcal{C}(\mathbf{X})$  and can be coupled so that  $\mathbf{X} \leftrightarrow \mathbf{B} \leftrightarrow (A_1, \dots, A_s)$  then it must be possible to couple  $\mathbf{U}_{\mathbf{B}}$  and  $(U_{A_1}, \dots, U_{A_s})$  to  $(\mathbf{X}, \mathbf{V})$  such that  $\mathbf{X} \leftrightarrow (\mathbf{V}, \mathbf{U}_{\mathbf{B}}) \leftrightarrow (U_{A_1}, \dots, U_{A_s})$ .
- 3) There exist functions  $g_1(\mathbf{V}, \mathbf{U}_{\mathbf{Y}_1}, \mathbf{Y}_1), \dots, g_m(\mathbf{V}, \mathbf{U}_{\mathbf{Y}_m}, \mathbf{Y}_m)$  such that  $E[d_i(\mathbf{X}, g_i(\mathbf{V}, \mathbf{U}_{\mathbf{Y}_i}, \mathbf{Y}_i))] \leq D_i + \epsilon 1$ , for all  $i \in [m]$ .

*Proof:* Let  $R$  be a  $\mathbf{D}$ -achievable rate,  $\epsilon > 0$ ,  $p(\mathbf{x})$  be given and  $p(\mathbf{y}_i|\mathbf{x})$ ,  $i \in [m]$  be fixed. Then there exists an  $(n, M, \mathbf{D} + \epsilon 1)$  code for some  $n$  such that  $H(I_0) \leq n(R + \epsilon)$ , where  $I_0$  is the output of the encoder. Also, let  $K(\mathbf{A}) = \frac{I(\mathbf{X}^n; I_0|\mathbf{A}^n)}{n}$ , where  $\mathbf{A}$  is a random vector with pmf  $\sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{a}|\mathbf{x})p(\mathbf{x})$ , i.e.,  $\mathbf{A} \in \mathcal{C}(\mathbf{X})$ . We call such  $\mathbf{A}$  *generalized side information*. Lastly, let  $\mathbf{V}'_i = I_0$ ,  $\mathbf{U}'_{\mathbf{A}i} = (\mathbf{A}_i^-, \mathbf{A}_i^+)$ , where  $\mathbf{A}_i^- = (\mathbf{A}_1, \dots, \mathbf{A}_{i-1})$  and  $\mathbf{A}_i^+ = (\mathbf{A}_{i+1}, \dots, \mathbf{A}_n)$  for  $i \in [n]$ , and let  $T$  denote a random variable that is uniformly distributed on  $[n]$  such that it is independent of the source  $\mathbf{X}$ , all generalized side information  $\mathbf{A}$ ,  $\mathbf{U}'_{\mathbf{A}i}$ , and  $\mathbf{V}'_i$ . Define  $\mathbf{U}_{\mathbf{A}} = (\mathbf{U}'_{\mathbf{A}}, T)$ ,  $\mathbf{V} = (\mathbf{V}', T)$ . Note that we have

$$R + \epsilon \geq K(\emptyset).$$

Also, we can write  $I(\mathbf{X}^n; I_0|\mathbf{X}^n) = 0$  and  $I(\mathbf{X}^n; I_0|\mathbf{A}^n) \geq 0$ , for all  $\mathbf{A}$ , giving the (initialize) and (non-negativity) conditions in the LP.

Let  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$ . For any such  $\mathbf{A}$  and  $\mathbf{B}$  we can write  $n(K(\mathbf{A}) - K(\mathbf{B}))$  as

$$\begin{aligned}
 I(\mathbf{B}^n; I_0|\mathbf{A}^n) &= \sum_{i=1}^n I(\mathbf{B}_i; I_0, \mathbf{B}_i^-, \mathbf{A}_i^+|\mathbf{A}_i) \\
 &\leq \sum_{i=1}^n I(\mathbf{B}_i; I_0, \mathbf{B}_i^-, \mathbf{B}_i^+|\mathbf{A}_i) \\
 &= \sum_{i=1}^n I(\mathbf{B}_i; \mathbf{V}'_i, \mathbf{U}'_{\mathbf{B}i}|\mathbf{A}_i).
 \end{aligned}$$

Since  $T$  is independent of  $\mathbf{X}, \mathbf{V}'$ , all generalized side information  $\mathbf{A}$  and all  $\mathbf{U}'_{\mathbf{B}}$ , we can write

$$\begin{aligned}
 n(K(\mathbf{A}) - K(\mathbf{B})) &\leq \sum_{i=1}^n I(\mathbf{B}_i; \mathbf{V}'_i, \mathbf{U}'_{\mathbf{B}i}|\mathbf{A}_i, T = i) \\
 &= nI(\mathbf{B}; \mathbf{V}', \mathbf{U}'_{\mathbf{B}}|\mathbf{A}, T) \\
 &= nI(\mathbf{B}; \mathbf{V}, \mathbf{U}_{\mathbf{B}}|\mathbf{A}),
 \end{aligned}$$

which gives the (slope) constraints in the LP.

Let  $(\mathbf{B} \succeq \mathbf{A}|\mathbf{C})$ . Then for each such coupling of  $(\mathbf{B}, \mathbf{A}, \mathbf{C})$ ,  $n(K((\mathbf{A}, \mathbf{C})) - K((\mathbf{B}, \mathbf{C})))$  is equal to

$$H(I_0|\mathbf{A}^n, \mathbf{C}^n) - H(I_0|\mathbf{B}^n, \mathbf{C}^n) \geq 0, \text{ by [5, Lemma 1],}$$

giving the (monotonicity) constraints in the LP.

Now we obtain the (monotonicity+) conditions in the LP. Let  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$ . By utilizing the chain rule again, we can write  $n(K(\mathbf{A}) - K(\mathbf{B}))$  as

$$\begin{aligned}
 I(\mathbf{B}^n; I_0|\mathbf{A}^n) &\geq \sum_{i=1}^n I(\mathbf{B}_i; I_0, \mathbf{A}_i^-, \mathbf{A}_i^+|\mathbf{A}_i) \\
 &= \sum_{i=1}^n I(\mathbf{B}_i; \mathbf{V}'_i, \mathbf{U}'_{\mathbf{A}i}|\mathbf{A}_i) \\
 &= nI(\mathbf{B}; \mathbf{V}, \mathbf{U}_{\mathbf{A}}|\mathbf{A}),
 \end{aligned} \tag{14}$$

giving (monotonicity+) conditions.

Let  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  be such that  $\mathbf{A} \leftrightarrow \mathbf{C} \leftrightarrow \mathbf{B}$  and  $\mathbf{C} = f_1(\mathbf{A})$  for some deterministic mapping  $f_1(\cdot)$ . By the chain rule,  $n(K(\mathbf{A}) + K(\mathbf{B}))$  is equal to

$$\begin{aligned} & I(\mathbf{B}^n; I_0 | \mathbf{A}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n, \mathbf{A}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n) \\ & \geq I(\mathbf{B}^n; I_0 | \mathbf{C}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n, \mathbf{A}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n) \\ & \geq I(\mathbf{B}^n; I_0 | \mathbf{C}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n, \mathbf{A}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n, \mathbf{C}^n) \\ & = I(\mathbf{X}^n, \mathbf{B}^n; I_0 | \mathbf{C}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n, \mathbf{A}^n) \\ & = I(\mathbf{X}^n; I_0 | \mathbf{C}^n) + I(\mathbf{X}^n; I_0 | \mathbf{B}^n, \mathbf{A}^n). \end{aligned}$$

By setting  $\mathbf{C} = f_2(\mathbf{B})$  and swapping the role of  $\mathbf{A}$  and  $\mathbf{B}$  in the procedure above, we get the (*submodularity*) conditions.

Now we find the properties of  $\mathbf{V}$  and  $\mathbf{U}_{\mathbf{A}}$  that give us the conditions 1)–3) in Theorem 3 and the Markov chain property in Definition 5. Let  $\mathbf{A}, \bar{\mathbf{A}} = (A_1, \dots, A_s)$  for some  $s$  be such that  $\mathbf{X} \leftrightarrow \mathbf{A} \leftrightarrow (A_1, \dots, A_s)$ . Firstly, since any set of  $\mathbf{U}'_{\mathbf{A}_i}$  is independent of  $\mathbf{X}_i$  and of any set of generalized side information  $A_i$ 's, all  $\mathbf{U}_{\mathbf{A}}$ 's are independent of  $\mathbf{X}$  and all  $\mathbf{A}$ 's. Secondly, note that  $\mathbf{X}_i \leftrightarrow (\mathbf{V}'_i, \mathbf{U}'_{\mathbf{A}_i}) \leftrightarrow (\mathbf{V}'_i, \mathbf{U}'_{A_1 i}, \dots, \mathbf{U}'_{A_s i})$  since

$$\begin{aligned} H(\mathbf{V}'_i, \mathbf{U}'_{\bar{\mathbf{A}}_i} | \mathbf{V}'_i, \mathbf{U}'_{\mathbf{A}_i}, \mathbf{X}_i) &= H(\bar{\mathbf{A}}_i^-, \bar{\mathbf{A}}_i^+ | I_0, \mathbf{A}_i^-, \mathbf{A}_i^+, \mathbf{X}_i) \\ &= H(\bar{\mathbf{A}}_i^-, \bar{\mathbf{A}}_i^+ | I_0, \mathbf{A}_i^-, \mathbf{A}_i^+). \end{aligned}$$

Then  $\mathbf{X} \leftrightarrow (\mathbf{V}', \mathbf{U}'_{\mathbf{A}}) \leftrightarrow (\mathbf{V}', \mathbf{U}'_{A_1}, \dots, \mathbf{U}'_{A_s})$  implies  $\mathbf{X} \leftrightarrow (\mathbf{V}, \mathbf{U}_{\mathbf{A}}) \leftrightarrow (\mathbf{V}, U_{A_1}, \dots, U_{A_s})$ . Furthermore, given  $(\mathbf{V}, \mathbf{U}_{\mathbf{Y}_i})$  and  $\mathbf{Y}_i, i \in [m]$ , decoder  $i$  can reconstruct the source subject to its own distortion constraint. Lastly,  $(\mathbf{V}, \mathbf{U}_{\mathbf{A}}) \leftrightarrow \mathbf{X} \leftrightarrow (\mathbf{A}, \mathbf{B})$  since  $(\mathbf{V}'_i, \mathbf{U}'_{\mathbf{A}_i}) \leftrightarrow \mathbf{X}_i \leftrightarrow (\mathbf{A}_i, \mathbf{B}_i)$  for all  $i \in [n]$ . ■

We can interpret  $K(\mathbf{A})$  in the LP as the amount of information that a hypothetical decoder with side information  $\mathbf{A}$  receives about  $\mathbf{X}$  from the broadcasted message. We can also view  $\mathbf{U}_{\mathbf{A}}$  as a quantized representation of the source that the hypothetical decoder can extract from the message with the help of its side information  $\mathbf{A}$  and  $\mathbf{V}$  as a common message to all decoders.

The (*submodularity*) condition is so named for the following reason. Let  $\mathbf{X} = (X_1, \dots, X_k)$ , where  $X_i$ 's are all independent random variables and let  $\mathbf{A} \subseteq \mathbf{X}, \mathbf{B} \subseteq \mathbf{X}$ <sup>2</sup>. Then we can write the (*submodularity*) condition for such  $\mathbf{A}$  and  $\mathbf{B}$  as  $K(\mathbf{A}) + K(\mathbf{B}) \geq K(\mathbf{A} \cap \mathbf{B}) + K(\mathbf{A} \cup \mathbf{B})$ . The LP lower bound was inspired by a similar lower bound for the special case of index coding [10]. That lower bound does not require auxiliary random variables, and it is expressed in terms of entropy instead of mutual information.

**Remark 4.** Evidently other conditions that  $I(\mathbf{X}^n; I_0 | \mathbf{A}^n)/n$  must satisfy in the context of this problem can be incorporated into the bound as desired.

**Remark 5.** The lower bound in Theorem 3 can be generalized to continuous sources with well-behaved distortion constraints such as Gaussian sources subject to component-wise mean square error (MSE) distortion constraints.

The lower bound in Theorem 3 is not evidently computable, since the infimum over  $K(\cdot)$  is subject to a continuum of constraints and there are no cardinality bounds on the  $V$  and  $U$  variables. We next provide a weakened lower bound that is computable. For this we need the following notation.

**Notation 8.** Let  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$  and  $\mathbf{D}_{\mathbf{A}} = \{\mathbf{D}_i | \mathbf{Y}_i \leftrightarrow \mathbf{A} \leftrightarrow \mathbf{X}\}$ . Then  $R(\mathbf{D}_{\mathbf{A}})$  denotes the result of the following optimization problem :

$$\min_{C_{\mathbf{A}}} I(\mathbf{B}; \mathbf{V} | \mathbf{A})$$

where

$C_{\mathbf{A}} : \mathbf{V} \in \mathcal{C}(\mathbf{X})$  such that

there exists functions  $g_i(\mathbf{V}, \mathbf{Y}_i)$  such that  $E[d_i(\mathbf{X}, g_i(\mathbf{V}, \mathbf{Y}_i))] \leq \mathbf{D}_i$  for all  $\mathbf{D}_i \in \mathbf{D}_{\mathbf{A}}$ .

**Theorem 4.** Let  $S_{\mathbf{A}}$  be a finite set of generalized side information variables  $\mathbf{A} \in \mathcal{C}(\mathbf{X})$  and consider the function  $K(\cdot)$  over the elements of  $S_{\mathbf{A}}$ . For any  $\epsilon > 0$ ,  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  is lower bounded by  $R'_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  where  $R'_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  is equal to

$$\inf K(\emptyset) - \epsilon, \quad (15)$$

where the infimum is over all  $K(\cdot) : S_{\mathbf{A}} \rightarrow [0, \infty)$  such that

$K(\mathbf{X}) = 0$  (initialize)

$K(\mathbf{A}) \geq 0$ , for all  $\mathbf{A}$  (non-negativity)

$K((\mathbf{A}, \mathbf{C})) \geq K((\mathbf{B}, \mathbf{C}))$ , for all  $(\mathbf{A}, \mathbf{B}) : (\mathbf{B} \succeq \mathbf{A} | \mathbf{C})$  (monotonicity)

$K(\mathbf{A}) \geq K(\mathbf{B}) + R(\mathbf{D}_{\mathbf{A}} + \epsilon \mathbf{1})$ , for all  $(\mathbf{A}, \mathbf{B}) : \mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$  (monotonicity+)

<sup>2</sup>Although  $\mathbf{X}$  is a vector, we can view it as an ordered set which also induces an ordered set structure on the subsets. Hence, we can use the set notation whenever it is convenient.

$K(\mathbf{A}) + K(\mathbf{B}) \geq K(\mathbf{C}) + K((\mathbf{A}, \mathbf{B}))$ , for all  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$  couplings :  $\mathbf{B} \leftrightarrow \mathbf{C} \leftrightarrow \mathbf{A}$  and  $\mathbf{C} = f_1(\mathbf{A})$  or  $\mathbf{C} = f_2(\mathbf{B})$  (submodularity)

*Proof of Theorem 4:* Let  $\epsilon > 0$ ,  $\mathbf{V} \in \mathcal{C}(\mathbf{X})$ , and  $\mathbf{U} : \mathcal{C}(\mathbf{X}) \rightarrow \mathcal{C}(\mathbf{X}, \mathbf{V})$  satisfying the conditions 1)–3) in Theorem 3 be given. Also, let LP1 be the linear program in Table I when  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  are in  $S_{\mathbf{A}}$  and let the solution of LP1 be denoted by  $\bar{R}_{lb}^{LP}(\epsilon)$ . Then  $R_{lb}^{LP}(\epsilon)$  in Theorem 3 is lower bounded by  $\bar{R}_{lb}^{LP}(\epsilon)$ . Therefore it is enough to show  $\bar{R}_{lb}^{LP}(\epsilon) \geq R'_{lb}(\mathbf{D} + \epsilon \mathbf{1})$ . Note that the constraints in LP1 and the LP in Theorem 4, denoted by LP2, are the same except the *monotonicity+* condition is different and there is no *slope* condition in LP2. But for any  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$  the *monotonicity+* condition in LP1 implies the *monotonicity+* condition in LP2 since  $I(\mathbf{B}; \mathbf{V}, \mathbf{U}_{\mathbf{A}} | \mathbf{A}) \geq R(\mathbf{D}_{\mathbf{A}} + \epsilon \mathbf{1})$  by condition 2) and 3) in Theorem 3. Hence,  $\bar{R}_{lb}^{LP}(\epsilon) \geq R'_{lb}(\mathbf{D} + \epsilon \mathbf{1})$ . ■

Note that  $R'_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  is computable since we have a finite number of constraints in the LP and each  $R(\mathbf{D}_{\mathbf{A}})$  can be computed by finding a cardinality constraint on the auxiliary random variable  $\mathbf{V}$  using standard techniques [14].

## V. COMPARISON WITH OTHER BOUNDS

### A. Upper Bound

Although there are achievable schemes for various forms of rate-distortion with side information (e.g. [4], [11], [12], [3]), most are for special cases of the problem. The two exceptions, both of which purport to provide achievable schemes for the general problem considered here, are Heegard and Berger [3] and Timo *et al.* [12]. Heegard and Berger's achievable rate was shown to be incorrect via a counterexample by Timo *et al.*, who also provided a corrected scheme. In fact, the proof of Timo *et al.*'s achievable result contains an error that is similar to that of Heegard and Berger. To see this, let us state Timo *et al.*'s achievable result.<sup>3</sup>

**Notation 9.**  $\bar{v} = \mathcal{S}_1, \dots, \mathcal{S}_{2^m-1}$  denotes an ordered list of all possible nonempty subsets of  $[m]$ , where each  $\mathcal{S}_i$  denotes a different subset such that  $|\mathcal{S}_i| \geq |\mathcal{S}_j|$  for all  $i < j$ .  $\bar{\mathcal{V}}$  denotes the set of all possible such  $\bar{v}$ .

**Notation 10.**

$$\begin{aligned} U_{\mathcal{S}_j}^{-'} &= \{U_{\mathcal{S}_i} \in \mathcal{U} \mid i < j, \mathcal{S}_i \not\supseteq \mathcal{S}_j\}, \\ U_{\mathcal{S}_j}^{\supseteq} &= \{U_{\mathcal{S}_i} \in \mathcal{U} \mid \mathcal{S}_i \supset \mathcal{S}_j\}, \\ U_{\mathcal{S}_j}^{+} &= \{U_{\mathcal{S}_k} \in \mathcal{U} \mid k > j, \mathcal{S}_k \cap \mathcal{S}_j \neq \emptyset\}, \\ U_{\mathcal{S}_j}^{\dagger} &= \left\{U_{\mathcal{S}_i} \in U_{\mathcal{S}_j}^{-'} \mid \begin{array}{l} \exists U_{\mathcal{S}_k} \in U_{\mathcal{S}_j}^{+} \\ \mathcal{S}_i \cap \mathcal{S}_k \neq \emptyset \end{array} \right\}, \text{ and} \\ U_{\mathcal{S}_j, l}^{\dagger} &= \{U_{\mathcal{S}_i} \in U_{\mathcal{S}_j}^{\dagger} : \mathcal{S}_i \ni l\} \text{ when } l \in \mathcal{S}_j. \end{aligned}$$

**Claim 1** (Theorem 2, [12]). *The rate-distortion function  $R(\mathbf{D})$  is upper bounded by*

$$R_{ach}^T(\mathbf{D}) = \min_{\bar{v} \in \bar{\mathcal{V}}} \inf_{C_{ach, \bar{v}}(\mathbf{D})} \inf_{C_T^{LP}} \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j}, \quad (16)$$

where  $C_{ach, \bar{v}}(\mathbf{D})$  is as in Theorem 1 and

$$\begin{aligned} C_T^{LP} : 1) & R_{\mathcal{S}_j} \geq 0, R'_{\mathcal{S}_j} \geq 0 \text{ for all } j \in [2^m - 1] \\ 2) & R_{\mathcal{S}_j} \geq I(X, U_{\mathcal{S}_j}^{\dagger}, U_{\mathcal{S}_j}^{\supseteq}; U_{\mathcal{S}_j}) - R'_{\mathcal{S}_j} \text{ for all } j \in [2^m - 1] \\ 3) & R'_{\mathcal{S}_j} \leq \min_{l \in \mathcal{S}_j} I(U_{\mathcal{S}_j}; U_{\mathcal{S}_j, l}^{\dagger}, U_{\mathcal{S}_j}^{\supseteq}, Y_l) \text{ for all } j \in [2^m - 1]. \end{aligned}$$

The proof given by Timo *et al.* proceeds as follows. Let  $\bar{v} \in \bar{\mathcal{V}}$  be given. The codebook generation is the same as in the proof of Theorem 1. Encoding is almost the same except that at each stage  $j$ , we select a codeword that is jointly typical with only those already-selected codewords that correspond to the messages  $U_{\mathcal{S}_j}^{\dagger}, U_{\mathcal{S}_j}^{\supseteq}$  and the source, instead of messages  $U_{\mathcal{S}_j}^{-}$  and the source as in Theorem 1. This creates an issue, however, because if the encoding proceeds in this fashion then there is no guarantee that the variables  $U_{\mathcal{S}_j}^{\dagger}, U_{\mathcal{S}_j}^{\supseteq}$  are themselves jointly typical.

To illustrate this, consider the case in which there are six decoders and suppose that  $\bar{v} = [6], \dots, \{1, 2\}, \{5, 6\}, \{3, 4\}, \{2, 3\}, \{4, 5\}, \{6\}, \{5\}, \{4\}, \{3\}, \{2\}, \{1\}$ . Choose  $\mathcal{U}$  such that all  $U_{\mathcal{S}_j} = \emptyset$  except  $U_{\{i, i+1\}}$ , for  $i \in [5]$ . Then the encoding order of the nontrivial messages is  $(U_{\{1, 2\}}, U_{\{5, 6\}}, U_{\{3, 4\}}, U_{\{2, 3\}}, U_{\{4, 5\}})$ . When the message  $U_{\{3, 4\}}$  is encoded, the encoder selects a codeword that is jointly typical with the codewords related to messages

<sup>3</sup>This problem also afflicts Theorem 1 in Timo *et al.*, although we shall focus our discussion on Theorem 2 of that paper, which is simpler and directly comparable to Theorem 1 in the present paper.



$U_{\{3,4\}}^\dagger = (U_{\{1,2\}}, U_{\{5,6\}})$  and the source (note that  $U_{\{3,4\}}^\supset = \emptyset$ ). However, in previous stages  $U_{\{1,2\}}$  and  $U_{\{5,6\}}$  were not selected in a way that guarantees that they are jointly typical, since  $U_{\{1,2\}} \notin \{U_{\{5,6\}}^\dagger \cup U_{\{5,6\}}^\supset\}$  and  $U_{\{5,6\}} \notin \{U_{\{1,2\}}^\dagger \cup U_{\{1,2\}}^\supset\}$ . The rate analysis in Timo *et al.*, specifically the use of Lemma 3 in that paper, presumes that the codewords corresponding to  $U_{\{1,2\}}$  and  $U_{\{5,6\}}$  are jointly typical when the codeword for  $U_{\{3,4\}}$  is chosen. This error is similar to the one in Heegard and Berger [3].<sup>4</sup> For the two-decoder case, this issue does not arise, and the Timo *et al.* rate is indeed achievable, as is that of Heegard and Berger.

This error could be fixed in several ways. Our scheme in Theorem 1 avoids this issue by requiring that each codeword be jointly typical with all of the previously-selected codewords. If a certain pair of auxiliary random variables never appear together in any of the mutual information expressions, then one can impose a conditional independence condition between them without loss of generality, which is tantamount, from a rate perspective, to not requiring that they be chosen in a way that ensures their joint typicality.

Our scheme in Theorem 1 differs from the achievable scheme in [12] in two other respects as well. We do not require that the sets in  $v$  be ordered so that their cardinalities are nonincreasing. Arguably the most notable difference is in the decoding. While in [12], each decoder decodes its messages sequentially in the same order that they are encoded, in our scheme we apply simultaneous decoding, i.e., we decode all messages for decoder  $i$  together. We shall see later, when discussing the odd-cycle index coding problem in Section VI, that for a given class of auxiliary random variables, simultaneous decoding can yield a strict rate improvement.

We conclude this subsection by showing that for the two-decoder case in which Claim 1 is valid, the upper bound in [12] is no worse than that of Theorem 1.

**Lemma 1.** *When there are two decoders,  $R_{ach}(\mathbf{D})$  is upper bounded by*

$$R^T(\mathbf{D}) = \min_{C_{ach,v}(\mathbf{D})} \max_{i \in \{1,2\}} \{I(X; U_{\{1,2\}} | Y_i)\} + I(X; U_{\{1\}} | U_{\{1,2\}}, Y_1) + I(X; U_{\{2\}} | U_{\{1,2\}}, Y_2), \quad (17)$$

where  $C_{ach,v}(\mathbf{D})$  is in Theorem 1.

*Proof of Lemma 1:* Firstly notice that  $U_{\{1\}}$  and  $U_{\{2\}}$  never appear together on the right-hand side of (17). Hence without loss of optimality we can add the condition  $U_{\{1\}} \perp U_{\{2\}} | X, U_{\{1,2\}}$  to  $C_{ach,v}(\mathbf{D})$ . Let  $v = \{\{1,2\}, \{1\}, \{2\}\}$  and  $U_{\mathcal{J}_j} \in C_{ach,v}(\mathbf{D})$  with  $U_{\{1\}} \perp U_{\{2\}} | X, U_{\{1,2\}}$ . From the LP conditions in Theorem 1, we can write

$$R_{\{1,2\}} + R'_{\{1,2\}} \geq I(X; U_{\{1,2\}}) \quad (18)$$

$$R_{\{1\}} + R'_{\{1\}} \geq I(X, U_{\{1,2\}}; U_{\{1\}}) \quad (19)$$

$$R_{\{2\}} + R'_{\{2\}} \geq I(X, U_{\{1,2\}}, U_{\{1\}}; U_{\{2\}}) \quad (20)$$

$$R'_{\{i\}} \leq I(U_{\{i\}}; U_{\{1,2\}}, Y_i), \text{ for all } i \in \{1, 2\} \quad (21)$$

$$R'_{\{1,2\}} \leq \min_{i \in \{1,2\}} \{I(U_{\{1,2\}}; U_{\{i\}}, Y_i)\} \quad (22)$$

$$R'_{\{1,2\}} + R'_{\{i\}} \leq I(U_{\{1,2\}}; Y_i) + I(U_{\{i\}}; U_{\{1,2\}}, Y_i), \text{ for all } i \in \{1, 2\}. \quad (23)$$

Then  $R'_{\{1,2\}} = \min_{i \in \{1,2\}} \{I(U_{\{1,2\}}; Y_i)\}$ ,  $R'_{\{i\}} = I(U_{\{i\}}; U_{\{1,2\}}, Y_i)$ ,  $R_{\{1,2\}} + R'_{\{1,2\}} = I(X; U_{\{1,2\}})$ ,  $R_{\{1\}} + R'_{\{1\}} = I(X, U_{\{1,2\}}; U_{\{1\}})$ , and  $R_{\{2\}} + R'_{\{2\}} = I(X, U_{\{1,2\}}, U_{\{1\}}; U_{\{2\}})$  are feasible choices enabling us to upper bound  $\inf_{C_{ach}^{LP}} \sum_{j=1}^3 R_{\mathcal{J}_j}$  in (2) by

$$\max_{i \in \{1,2\}} \{I(X; U_{\{1,2\}} | Y_i)\} + I(X; U_{\{1\}} | U_{\{1,2\}}, Y_1) + I(X; U_{\{2\}} | U_{\{1,2\}}, Y_2) + I(U_{\{1\}}; U_{\{2\}} | X, U_{\{1,2\}}), \quad (24)$$

which is equal to the mutual information expression in Lemma 1 when  $U_{\{1\}} \perp U_{\{2\}} | X, U_{\{1,2\}}$ . Therefore,  $R^T(\mathbf{D}) \geq R_{ach}(\mathbf{D})$ . ■

## B. Lower Bounds

1) *minimax-type Lower Bound:* First we compare the general lower bound,  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$ , with the minimax version of the lower bound in [11]. For completeness, we state the minimax version of the theorem below.

**Theorem 5.** *Let the pmf's  $p(x, y_i)$  for all  $i \in [m]$  be given. Then  $R(\mathbf{D})$  is lower bounded by*

$$R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) = \sup_{\bar{P}} \inf_{\bar{C}} \bar{R}_{lb} - \epsilon, \quad (25)$$

$$\begin{aligned} \text{where } \bar{R}_{lb} = & \max_{\sigma} [I(X; V, U_{Y_{\sigma(1)}} | Y_{\sigma(1)}) \\ & + I(X; U_{Y_{\sigma(2)}} | V, U_{Y_{\sigma(1)}}, Y_{\sigma(1)}, Y_{\sigma(2)}) + \cdots \\ & + I(X; U_{Y_{\sigma(m)}} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(m-1)}}), Y)] \end{aligned} \quad (26)$$

<sup>4</sup>Unlike the Heegard-Berger result, however, the rate promised by Timo *et al.*'s achievable result is not known to be unachievable in general at this point.

$Y = (Y_{\sigma(1)}, \dots, Y_{\sigma(m)}),$  and

1)  $\bar{P} = \{p(x, y_1, \dots, y_m) | \sum_{y_j: j \neq i} p(x, y_1, \dots, y_m) = p(x, y_i), \forall i \in [m]\}.$

2)  $\bar{C}$  denotes the set of  $(V, U_{Y_1}, \dots, U_{Y_m})$  jointly distributed with  $X, Y_1, \dots, Y_m$  such that  $(Y_1, \dots, Y_m) \leftrightarrow X \leftrightarrow (V, U_{Y_1}, \dots, U_{Y_m})$  and there exists functions  $g_1, \dots, g_m$  with the property that  $\mathbb{E}[d_i(X, g_i(V, U_{Y_i}, Y_i))] \leq D_i + \epsilon, \forall i \in [m].$

3)  $\sigma(\cdot)$  denotes a permutation on integers  $[m].$

The minimax lower bound in Theorem 5 is the state-of-the-art for the general rate-distortion problem with side information at multiple decoders. Note that in Theorem 5, one can absorb  $V$  into  $U_{Y_i}, i \in [m]$  without loss of optimality. For the ease of comparison with  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  we leave it as a separate variable, however.

**Theorem 6.**  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}),$  where  $\epsilon > 0.$

*Proof:* Consider  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1}).$  Note that the LP constraints of  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  apply to all choices of the relevant random variables. Hence we can write

$$R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq \sup_{\bar{P}} \inf_{V \in \mathcal{C}(X)} \inf_{U: \mathcal{C}(X) \rightarrow \mathcal{C}(X, V)} R_{lb}^{LP}(\epsilon) \quad (27)$$

where  $\bar{P}$  is as in Theorem 5, and  $V$  and  $U$  in the infima satisfy the conditions 1)–3) in Theorem 3 for a fixed coupling of the random variables. Now we find a lower bound to the quantity  $R_{lb}^{LP}(\epsilon)$  in (27) by utilizing the *monotonicity* and *monotonicity+* constraints of the LP in Table I. We can write the following series of inequalities:

$$K(\emptyset) \geq K(Y_1) \text{ by (monotonicity)} \quad (28)$$

$$K(Y_1) \geq K(Y_1, Y_2) + I(Y_2; V, U_{Y_1} | Y_1) \quad (29)$$

$\vdots$

$$K(Y_1, \dots, Y_m) \geq K(Y_1, \dots, Y_m, X) + I(X; V, U_{Y_1}, \dots, U_{Y_m} | Y_1, \dots, Y_m) \quad (30)$$

$$K(Y_1, \dots, Y_m, X) = 0. \quad (31)$$

where (29) is from *monotonicity+* and (30) is from *monotonicity+* and  $(Y_1, \dots, Y_m) \leftrightarrow X \leftrightarrow (V, U_{Y_1}, \dots, U_{Y_m}) \leftrightarrow (V, U_{Y_1}, \dots, U_{Y_m}).$  If we add all these inequalities side-by-side we obtain

$$\begin{aligned} K(\emptyset) &\geq I(Y_2; V, U_{Y_1} | Y_1) + \dots \\ &\quad + I(Y_m; V, U_{Y_1}, \dots, U_{Y_{m-1}} | Y_1, \dots, Y_{m-1}) \\ &\quad + I(X; V, U_{Y_1}, \dots, U_{Y_m} | Y_1, \dots, Y_m). \end{aligned} \quad (32)$$

By applying a series of chain rules and combining terms, we can write the right-hand side of (32) as

$$\begin{aligned} &I(X; V, U_{Y_1} | Y_1) + \dots + I(X; U_{Y_2} | V, U_{Y_1}, Y_1, Y_2) \\ &\quad + I(X; U_{Y_m} | V, U_{Y_1}, \dots, U_{Y_{m-1}}, Y_1, \dots, Y_m). \end{aligned}$$

Let us define

$$\begin{aligned} \Gamma_k &= \sum_{i=2}^k I(Y_i; V, U_{Y_1}, \dots, U_{Y_{i-1}} | Y_1, \dots, Y_{i-1}) \\ &\quad + I(X; V, U_{Y_1}, \dots, U_{Y_k} | Y_1, \dots, Y_k) \\ &\quad + \sum_{i=k+1}^m I(X; U_{Y_i} | V, U_{Y_1}, \dots, U_{Y_{i-1}}, Y_1, \dots, Y_i) \end{aligned}$$

for  $k \in [m]$  where “empty” sums are zero. Note that  $\Gamma_m$  is equal to the right-hand side of (32). One can show that  $\Gamma_1 = \Gamma_2 = \dots = \Gamma_m.$  Hence  $K(\emptyset) \geq \Gamma_1.$

Also since there are  $m$  decoders, we can get  $m!$  lower bounds on  $K(\emptyset)$  by considering all possible permutations on integers  $[m].$  Hence, we have  $K(\emptyset) \geq \bar{R}_{lb}.$  From (27) we can write

$$R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq \sup_{\bar{P}} \inf_{V \in \mathcal{C}(X)} \inf_{U: \mathcal{C}(X) \rightarrow \mathcal{C}(X, V)} \bar{R}_{lb} - \epsilon \quad (33)$$

$$\geq \sup_{\bar{P}} \inf_{\bar{C}} \bar{R}_{lb} - \epsilon, \quad (34)$$

where  $\bar{C}$  is as in Theorem 5. Lastly, we have (34) since each feasible set of random variables in the infima in (33) is also feasible for  $\bar{C}.$  Hence,  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}).$  ■

TABLE II  
LP BOUND FOR INDEX CODING PROBLEM

$\min \widehat{K}(\emptyset)$  subject to  
 $\widehat{K}(\mathbf{X}) \geq |\mathbf{X}|$  (*initialize*)  
 $\widehat{K}(\mathbf{A}) + |\mathbf{B} \setminus \mathbf{A}| \geq \widehat{K}(\mathbf{B})$ , for all  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$  (*slope*)  
 $\widehat{K}(\mathbf{B}) \geq \widehat{K}(\mathbf{A})$ , for all  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$  (*monotonicity*)  
 $\widehat{K}(\mathbf{A}) = \widehat{K}(\mathbf{B})$ , for all  $\mathbf{A}, \mathbf{B} \subseteq \mathbf{X} : \mathbf{A} \sim \mathbf{B}$  (*decode*)  
 $\widehat{K}(\mathbf{A}) + \widehat{K}(\mathbf{B}) \geq \widehat{K}(\mathbf{A} \cap \mathbf{B}) + \widehat{K}(\mathbf{A} \cup \mathbf{B})$ ,  
 for all  $\mathbf{A}, \mathbf{B} \subseteq \mathbf{X}$  (*submodularity*).

2) *LP Lower Bound for the Index Coding Problem:* We next compare the general lower bound,  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  with the linear programming lower bound in [10] for the index coding problem [9]. In the index coding problem, the source  $\mathbf{X} = (X_1, \dots, X_k)$  is such that  $X_i$ ,  $i \in [k]$  are independent and identically distributed (i.i.d.) Bernoulli  $(\frac{1}{2})$  random variables and each side information  $\mathbf{Y}_i$  at decoder  $i$  is an arbitrary subset<sup>5</sup> of the source  $\mathbf{X}$ . Each decoder  $i$  wishes to reconstruct an arbitrary subset of the source,  $\widehat{\mathbf{X}}_i \subseteq \mathbf{X} \setminus \mathbf{Y}_i$ . The reconstructions can either be required to be zero error [10] or such that the block error probability vanishes [11]. Both formulations are more stringent than considering the problem with Hamming distortion in the limit in which the distortion goes to zero, so  $R_{lb}(\epsilon \mathbf{1})$  is a valid lower bound to the index coding problem in all three cases.

We first state the LP lower bound in [10], originally stated for the zero-error form of the problem. For completeness, we need the following notation.

**Notation 11.**  $\mathbf{A} \rightsquigarrow \mathbf{B}$  denotes “ $\mathbf{A}$  decodes  $\mathbf{B}$ ,” meaning that  $\mathbf{A} \subseteq \mathbf{B}$  and for every source component  $X_i \in \mathbf{B} \setminus \mathbf{A}$  there is a decoder  $j$  who reconstructs  $X_i$  and  $\mathbf{Y}_j \subseteq \mathbf{A}$ . Also  $S(\mathbf{A}) = \{X_i \mid \text{decoder } j \text{ reconstructs } X_i \in \mathbf{X} \text{ and } \mathbf{Y}_j \subseteq \mathbf{A}\}$ .

**Theorem 7** (LP lower bound [10]). *The optimal value for the linear program in Table II<sup>6</sup>, denoted by  $\widehat{R}_{lb}^{LP}$ , is a lower bound to the index coding problem.*

Now that we stated the LP lower bound in [10], we show that  $\lim_{\epsilon \rightarrow 0} R_{lb}(\epsilon \mathbf{1})$  is equal to this bound when we restrict the generalized side information,  $\mathbf{A}$ , in  $R_{lb}(\epsilon \mathbf{1})$  to be a subset of the source,  $\mathbf{X}$ . From now on we denote this weakened form of  $R_{lb}(\epsilon \mathbf{1})$  obtained by restricting the generalized side information to be a subset of the source by  $R_{lb}^I(\epsilon \mathbf{1})$ . The following two lemmas will be useful to prove that the weakened lower bound  $R_{lb}^I(\epsilon \mathbf{1})$  is equal to the LP lower bound in Theorem 7.

**Lemma 2.** *Without loss of optimality we can replace the (*initialize*) and (*slope*) conditions in the LP in Table II with*

$$\begin{aligned}
 \widehat{K}(\mathbf{X}) &= |\mathbf{X}| \text{ (initialize*)} \\
 \widehat{K}(\mathbf{A}) + |\mathbf{B} \setminus \{S(\mathbf{A}) \cup \mathbf{A}\}| &\geq \widehat{K}(\mathbf{B}), \text{ for all } \mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X} \text{ (slope*)},
 \end{aligned}$$

respectively.

*Proof:* First we show that without loss of optimality we can add the *initialize\** and *slope\** conditions to the LP in Table II. Since they are more stringent than *initialize* and *slope* conditions in Table II, the result then follows. We begin with *initialize\**. Let  $\widehat{K}(\mathbf{A})$ ,  $\mathbf{A} \subseteq \mathbf{X}$  be feasible for the LP in Table II such that  $\widehat{K}(\mathbf{X}) > |\mathbf{X}|$ . Then there exists  $\epsilon > 0$  such that  $\widehat{K}(\mathbf{X}) = |\mathbf{X}| + \epsilon$ . Note that  $\widehat{K}(\mathbf{A}) - \epsilon$ ,  $\mathbf{A} \subseteq \mathbf{X}$ , is also feasible for the LP in Table II giving a lower objective  $\widehat{K}(\emptyset) - \epsilon$ . Hence, without loss of optimality we can insert the *initialize\** condition into the LP in Table II. Now we show that the *slope* and *decode* conditions of the LP in Table II imply the *slope\** condition. Let  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$ . If  $\mathbf{B} \cap S(\mathbf{A}) = \emptyset$  then the *slope* and *slope\** conditions are equivalent. Otherwise, i.e., if  $\mathbf{B} \cap S(\mathbf{A}) = \mathbf{C} \neq \emptyset$ , then from the *decode* and *slope* conditions we have

$$\begin{aligned}
 \widehat{K}(\mathbf{C} \cup \mathbf{A}) &= \widehat{K}(\mathbf{A}), \\
 \widehat{K}(\mathbf{C} \cup \mathbf{A}) + |\mathbf{B} \setminus \{\mathbf{C} \cup \mathbf{A}\}| &\geq \widehat{K}(\mathbf{B})
 \end{aligned}$$

respectively. Since  $\mathbf{B} \setminus \{\mathbf{C} \cup \mathbf{A}\} = \mathbf{B} \setminus \{S(\mathbf{A}) \cup \mathbf{A}\}$ , the *decode* and *slope* conditions imply the *slope\** condition. ■

**Lemma 3.** *Let  $\epsilon > 0$  and  $\bar{R}_{lb}^{LP}(\epsilon \mathbf{1})$  be the optimal value of the LP in Table III. Then  $R_{lb}^I(\epsilon \mathbf{1}) \geq \bar{R}_{lb}^{LP}(\epsilon \mathbf{1})$  and  $\lim_{\epsilon \rightarrow 0} R_{lb}^I(\epsilon \mathbf{1}) = \bar{R}_{lb}^{LP}(\mathbf{0})$ .*

*Proof:* Since the random variables  $\mathbf{A}, \mathbf{B}$  in  $R_{lb}^I(\epsilon \mathbf{1})$  are such that  $\mathbf{A}, \mathbf{B} \subseteq \mathbf{X}$ , the Markov chain  $\mathbf{A} \leftrightarrow \mathbf{B} \leftrightarrow \mathbf{X}$  is equivalent to  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$ . Then the *slope* constraints of the LP in  $R_{lb}^I(\epsilon \mathbf{1})$  imply the *slope* constraints of  $\bar{R}_{lb}^{LP}(\epsilon \mathbf{1})$ , since  $H(\mathbf{B}|\mathbf{A}) \geq I(\mathbf{B}; \mathbf{V}, \mathbf{U}_B|\mathbf{A})$ . Furthermore, using Fano's inequality, it can be seen that the *monotonicity+* condition of the

<sup>5</sup>Although  $\mathbf{X}$  is a vector, we can view it as an ordered set which also induces an ordered set structure on the subsets. Hence, we can use the set notation whenever it is convenient.

<sup>6</sup>The statement of the result in [10] does not contain the (monotonicity) condition, although it is clear from the proof that it was intended to be included. The condition is present in the preprint version of the paper [15].

TABLE III  
RELAXATION OF THE LP IN TABLE I

$\min K(\emptyset) - \epsilon$  subject to  
 $K(\mathbf{X}) = 0$  (*initialize*)  
 $K(\mathbf{A}) \geq 0$ , for all  $\mathbf{A} \subseteq \mathbf{X}$  (*non-negativity*)  
 $K(\mathbf{B}) + H(\mathbf{B}|\mathbf{A}) \geq K(\mathbf{A})$ , for all  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$  (*slope*)  
 $K(\mathbf{A}) \geq K(\mathbf{B})$ , for all  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$  (*monotonicity*)  
 $K(\mathbf{A}) \geq K(\mathbf{B}) + H(\mathbf{B}|\mathbf{A}) - H(\mathbf{B}|S(\mathbf{A}), \mathbf{A}) - \epsilon \log |S(\mathbf{A})|$ , for all  $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{X}$  (*monotonicity+*)  
 $K(\mathbf{A}) + K(\mathbf{B}) \geq K(\mathbf{A} \cap \mathbf{B}) + K(\mathbf{A} \cup \mathbf{B})$ ,  
 for all  $\mathbf{A}, \mathbf{B} \subseteq \mathbf{X}$  (*submodularity*).

LP in  $R_{lb}^I(\epsilon \mathbf{1})$  gives the *monotonicity+* condition of  $\bar{R}_{lb}^{LP}(\epsilon \mathbf{1})$  and the rest of the conditions are the same. Hence, we have  $R_{lb}^I(\epsilon \mathbf{1}) \geq \bar{R}_{lb}^{LP}(\epsilon \mathbf{1})$ . Now we select  $\mathbf{V} = \mathbf{Z}$  where  $\mathbf{Z}$  is a vector of i.i.d. Bernoulli( $\frac{1}{2}$ ) bits of the same length as  $\mathbf{X}$ ,  $\mathbf{Z} \perp \mathbf{X}$ , and we select  $\mathbf{U}_A = (S(\mathbf{A}), \mathbf{A}) \oplus \mathbf{Z}$ ,<sup>7</sup>  $\mathbf{A} \subseteq \mathbf{X}$ . Note that this selection of  $\mathbf{V}$  and  $\mathbf{U}_A$  satisfy the conditions 1)–3) in Theorem 3. Then the solution of the resulting LP is equal to the LP in Table III where  $\epsilon \log |S(\mathbf{A})| = 0$ , giving  $\bar{R}_{lb}^{LP}(\mathbf{0}) - \epsilon \geq R_{lb}^I(\epsilon \mathbf{1})$ . Since  $\bar{R}_{lb}^{LP}(\epsilon \mathbf{1})$  is right-continuous at  $\epsilon = 0$  [13], letting  $\epsilon \rightarrow 0$  gives the result. ■

**Theorem 8.**  $\lim_{\epsilon \rightarrow 0} R_{lb}^I(\epsilon \mathbf{1}) = \hat{R}_{lb}^{LP}$ .

*Proof:* Let  $LP_1$  and  $LP_2$  denote the LPs in Theorem 7 and Table III with  $\epsilon = 0$ , respectively. By Lemma 2, without loss of optimality we can add the *initialize\** and *slope\** conditions in Lemma 2 to  $LP_1$  and consider  $LP_1$  of this form. Notice that  $\bar{R}_{lb}^{LP}(\mathbf{0})$  is the solution of  $LP_2$  and from Lemma 3,  $\lim_{\epsilon \rightarrow 0} R_{lb}^I(\epsilon \mathbf{1}) = \bar{R}_{lb}^{LP}(\mathbf{0})$ . Hence, it is enough to show that  $\hat{R}_{lb}^{LP} = \bar{R}_{lb}^{LP}(\mathbf{0})$ . We show this by reparametrizing  $LP_2$  in terms of  $\hat{K}(\mathbf{A})$  where  $\hat{K}(\mathbf{A}) = K(\mathbf{A}) + H(\mathbf{A})$ . Note that  $\hat{K}(\emptyset) = K(\emptyset)$ . Hence, the objective of  $LP_2$  is the same as the objective of  $LP_1$ . Now we show that the constraint set in  $LP_2$  and the constraint set in  $LP_1$  are the same. We can rewrite the *initialize* and *non-negativity* conditions of  $LP_2$  as

$$\hat{K}(\mathbf{X}) = H(\mathbf{X})$$

$\hat{K}(\mathbf{A}) \geq H(\mathbf{A})$  respectively. Together those two conditions are equivalent to the *initialize\** and *slope* conditions of  $LP_1$ .

When we rewrite the *slope* condition of  $LP_2$ , we get

$$\hat{K}(\mathbf{B}) \geq \hat{K}(\mathbf{A}), \text{ the monotonicity condition of } LP_1.$$

When we rewrite the *monotonicity* and *monotonicity+* conditions of  $LP_2$ , we get

$$\hat{K}(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \geq \hat{K}(\mathbf{B})$$

$\hat{K}(\mathbf{A}) + H(\mathbf{B}|S(\mathbf{A}), \mathbf{A}) \geq \hat{K}(\mathbf{B})$  respectively and they are equivalent to the *slope* and *slope\** conditions of  $LP_1$ .

Also, combining the *submodularity* condition of  $LP_2$  and  $H(\mathbf{A}) + H(\mathbf{B}) = H(\mathbf{B} \cap \mathbf{A}) + H(\mathbf{B} \cup \mathbf{A})$  we can get the same *submodularity* condition of  $LP_1$ .

Lastly, from the *monotonicity+* and *slope* conditions of  $LP_2$ , we can obtain  $K(\mathbf{A}) + H(\mathbf{A}) = K(\mathbf{B}) + H(\mathbf{B}|\mathbf{A}) + H(\mathbf{A})$  for all  $\mathbf{A} \rightsquigarrow \mathbf{B}$ , which is the *decode* condition of  $LP_1$ . Hence, each constraint (or combination of constraints) in  $LP_2$  corresponds to a constraint in  $LP_1$  and vice versa. Since the objectives of  $LP_1$  and  $LP_2$  are the same, we conclude that  $\hat{R}_{lb}^{LP} = \bar{R}_{lb}^{LP}(\mathbf{0})$ . ■

## VI. OPTIMALITY RESULTS

The LP upper and lower bounds are tight in several instances<sup>8</sup>. We begin with several classes of instances for which the rate-distortion function is already known, the last of which is the odd-cycle index coding problem, which can be considered as a special case of Heegard-Berger problem. We conclude this section by finding an explicit characterization of the rate-distortion function for a new “odd-cycle Gaussian problem” using the upper and lower bounds in Theorems 1 and 4, respectively.

### A. Rate-Distortion Function with Mismatched Side Information at Decoders [4]

In this problem, there is one encoder with source  $\mathbf{X} = (X_1, X_2)$  and two decoders with side information  $\mathbf{Y}_1 = (Y_{11}, Y_{12})$  and  $\mathbf{Y}_2 = (Y_{21}, Y_{22})$ , respectively. The source and side information satisfy the following relations

$$(X_1, Y_{11}, Y_{21}) \perp (X_2, Y_{12}, Y_{22}) \quad (35)$$

$$X_1 \leftrightarrow Y_{11} \leftrightarrow Y_{21} \text{ and } X_2 \leftrightarrow Y_{12} \leftrightarrow Y_{22} \quad (36)$$

and the reconstructions at the decoders,  $\hat{\mathbf{X}}_1 = (\hat{X}_{11}, \hat{X}_{12})$  and  $\hat{\mathbf{X}}_2 = (\hat{X}_{21}, \hat{X}_{22})$ , are such that

$$\mathbb{E}[d_{1i}(X_1, \hat{X}_{1i})] \leq D_{1i} \quad (37)$$

$$\mathbb{E}[d_{2i}(X_2, \hat{X}_{2i})] \leq D_{2i} \text{ for } i \in [2]. \quad (38)$$

<sup>7</sup>  $\mathbf{a} \oplus \mathbf{b}$  denotes componentwise exclusive-OR operation where the shorter vector is zero padded as necessary.

<sup>8</sup> In a recent work of Benammar *et al.* [6], the rate-distortion problem with two decoders having degraded reconstruction sets is considered and the corresponding rate-distortion function is characterized. The construction of auxiliary random variables in the converse result of Benammar *et al.* [6] is specific to that problem setting and at this point it is unclear whether the LP lower bound subsumes this converse result.

We denote the rate-distortion function of this problem as  $R^M(\mathbf{D})$ . Theorem 10 shows that the minimax lower bound in Theorem 5 is greater than or equal to  $R^M(\mathbf{D})$ , the rate-distortion function characterized by Watanabe [4]. Hence, it implies that the lower bounds in both Theorems 5 and 3 are tight for this problem.

**Theorem 9** ([4]). *The rate-distortion function,  $R^M(\mathbf{D})$ , equals*

$$\begin{aligned} R^M(\mathbf{D}) &= \min[\max\{R_1^M, R_2^M\}], \text{ where} \\ R_1^M &= I(X_1; W_1|Y_{11}) + I(X_2; W_2|Y_{12}) + I(X_1; U_1|Y_{11}, W_1) + I(X_2; U_2|Y_{22}, W_2) \\ R_2^M &= I(X_1; W_1|Y_{21}) + I(X_2; W_2|Y_{22}) + I(X_1; U_1|Y_{11}, W_1) + I(X_2; U_2|Y_{22}, W_2), \end{aligned}$$

and the minimization is taken over all auxiliary random variables  $W_1, W_2, U_1, U_2$  satisfying the following:

- 1)  $(W_i, U_i) \leftrightarrow X_i \leftrightarrow (Y_{1i}, Y_{2i})$  for  $i = 1, 2$ .
- 2)  $(W_1, U_1, X_1, Y_{11}, Y_{21})$  and  $(W_2, U_2, X_2, Y_{12}, Y_{22})$  are independent of each other.
- 3) There exist functions  $g_{11}(W_1, U_1, Y_{11}) = \hat{X}_{11}$ ,  $g_{12}(W_2, Y_{12}) = \hat{X}_{12}$ ,  $g_{21}(W_1, Y_{21}) = \hat{X}_{21}$ , and  $g_{22}(W_2, U_2, Y_{22}) = \hat{X}_{22}$  such that they satisfy (37) and (38).
- 4)  $|\mathcal{W}_i| \leq |\mathcal{X}_i| + 3$  and  $|\mathcal{U}_i| \leq |\mathcal{X}_i| \cdot (|\mathcal{X}_i| + 3) + 1$  for  $i = 1, 2$ , where  $\mathcal{W}_i$  and  $\mathcal{U}_i$  are alphabets of  $W_i$  and  $U_i$  respectively.

**Theorem 10.**  $\liminf_{\epsilon \rightarrow 0} R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) \geq R^M(\mathbf{D})$  and  $R_{ach}(\mathbf{D}) \leq R^M(\mathbf{D})$ .

*Proof:* We select the joint distribution of  $(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2)$  such that it satisfies (36). First we show  $\liminf_{\epsilon \rightarrow 0} R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) \geq R^M(\mathbf{D})$ . Let  $U_Y = (V, U_{Y_1})$  and  $U_Z = (V, U_{Y_2})$ . Then  $\bar{R}_{lb}$  in Theorem 5 can be written as<sup>9</sup>  $\bar{R}_{lb} = \max\{\bar{R}_{lb1}, \bar{R}_{lb2}\}$ ,

$$\begin{aligned} \bar{R}_{lb1} &= I(\mathbf{X}; U_Y | \mathbf{Y}_1) + I(\mathbf{X}; U_Z | U_Y, \mathbf{Y}_1, \mathbf{Y}_2) \\ \bar{R}_{lb2} &= I(\mathbf{X}; U_Z | \mathbf{Y}_2) + I(\mathbf{X}; U_Y | U_Z, \mathbf{Y}_1, \mathbf{Y}_2). \end{aligned}$$

By the chain rule and using (36),  $\bar{R}_{lb1}$  can be rewritten as

$$\begin{aligned} &I(X_2; U_Y, Y_{11} | Y_{12}) + I(X_1; U_Y, Y_{12}, X_2 | Y_{11}) + I(X_2; U_Z | U_Y, Y_{11}, Y_{22}) + I(X_1; U_Z | U_Y, Y_{22}, X_2, Y_{11}) \\ &\stackrel{a}{=} I(X_2; U_Y, Y_{11} | Y_{12}) + I(X_1; U_Y, Y_{22}, X_2 | Y_{11}) + I(X_2; U_Z | U_Y, Y_{11}, Y_{22}) + I(X_1; U_Z | U_Y, Y_{22}, X_2, Y_{11}) \\ &\stackrel{b}{=} I(X_2; U_Y, Y_{11} | Y_{12}) + I(X_1; U_Y, Y_{22}, X_2, U_Z | Y_{11}) + I(X_2; U_Z | U_Y, Y_{11}, Y_{22}) \\ &\geq I(X_2; U_Y, Y_{11} | Y_{12}) + I(X_1; U_Y, Y_{22}, U_Z | Y_{11}) + I(X_2; U_Z | U_Y, Y_{11}, Y_{22}), \end{aligned}$$

which equals  $I(X_2; W_2 | Y_{12}) + I(X_1; W_1, U_1 | Y_{11}) + I(X_2; U_2 | W_2, Y_{22}) = R_1^M$ , where  $W_2 = (V, U_{Y_1}, Y_{11})$ ,  $W_1 = (V, U_{Y_2}, Y_{22})$ ,  $U_1 = U_{Y_1}$  and  $U_2 = U_{Y_2}$ . Here

a: follows since  $I(X_1; Y_{12}, Y_{22} | X_2, U_Y, U_Z, Y_{11}) = 0$ .

b: follows by combining the second and last term.

Similarly, we can obtain  $\bar{R}_{lb2} \geq R_2^M$ .

Note that  $(U_Y, U_Z) \leftrightarrow (X_1, X_2) \leftrightarrow (Y_{11}, Y_{12}, Y_{21}, Y_{22})$  implies the first condition of the minimization in Theorem 9. Also, the distortion constraints in  $R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1})$  imply the third condition of the minimization with  $\epsilon$  added to distortion constraints in Theorem 9. Hence, we can write

$$R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) \geq \inf[\max\{R_1^M, R_2^M\}] - \epsilon, \quad (39)$$

where the minimization is over  $(W_1, U_1, W_2, U_2)$  satisfying the first and third conditions of the minimization in Theorem 9. Also, since  $(W_1, U_1)$  and  $(W_2, U_2)$  do not appear together, we can add the condition 2) in Theorem 9 to the minimization in (39). Lastly, cardinality bounds on  $(W_1, W_2, U_1, U_2)$  can be obtained as in  $R^M(\mathbf{D})$  and the right-hand side of (39) can be shown to be continuous in  $\epsilon$  using the same procedure as in [4].

It remains to show that  $R_{ach}(\mathbf{D}) \leq R^M(\mathbf{D})$ . In [4],  $R^T(\mathbf{D})$  in Lemma 1 is utilized to obtain  $R^M(\mathbf{D})$ . Hence, from Lemma 1, we have  $R_{ach}(\mathbf{D}) \leq R^M(\mathbf{D})$ .  $\blacksquare$

### B. Rate-Distortion Function with Conditionally Less Noisy Side Information [5]

There are two decoders, and the distortion measure at decoder 1,  $d_1(\cdot, \cdot)$ , is such that  $d_1(X, \hat{X}) = 0$  if  $\hat{X} = a(X)$  and  $d_1(X, \hat{X}) = 1$  otherwise, where  $a(X)$  is a deterministic map. Also the allowable distortion at decoder 1,  $D_1$ , is taken as zero. Timo *et al.* [5] show that their lower bound for this problem is tight if  $Y_2$  is conditionally less noisy than  $Y_1$ , i.e.,  $(Y_2 \succeq Y_1 | a(X))$ , and  $H(a(X) | Y_1) \geq H(a(X) | Y_2)$ . Although whether the minimax lower bound in Theorem 5 is tight for this problem is not known, the next theorem shows that  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  subsumes the lower bound in [5] when  $(Y_2 \succeq Y_1 | a(X))$ .

<sup>9</sup>Note that Theorem 5 can be applied to vector-valued sources and side information at the decoders.

**Theorem 11.**  $\liminf_{\epsilon \rightarrow 0} R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq R^{LN}(\mathbf{D})$  and  $R_{ach}(\mathbf{D}) \leq R^{LN}(\mathbf{D})$  where

$$R^{LN}(\mathbf{D}) = H(a(X)|Y_1) + \min_{\substack{W \leftrightarrow X \leftrightarrow (a(X), Y_2) \\ \mathbb{E}[d_2(X, g_2(W, a(X), Y_2))] \leq D_2 \\ |\mathcal{W}| \leq |\mathcal{X}| + 1,}} I(X; W|a(X), Y_2)$$

is the lower bound in [5, Lemma 5] when  $(Y_2 \succeq Y_1|a(X))$ .

*Proof:* We begin with showing  $\liminf_{\epsilon \rightarrow 0} R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq R^{LN}(\mathbf{D})$ . Similar to the proof of Theorem 5, first we consider  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$ . For a given  $\epsilon > 0$  we can write

$$R_{lb}(\mathbf{D} + \epsilon \mathbf{1}) \geq \inf_{V \in \mathcal{C}(X)} \inf_{U: \mathcal{C}(X) \rightarrow \mathcal{C}(X, V)} R_{lb}^{LP}(\epsilon) \quad (40)$$

where the  $LP$  constraints on the random variables  $(X, a(X), Y_1, Y_2)$  are as in the problem description. Now we find a lower bound to  $R_{lb}^{LP}(\epsilon)$  in (40) by utilizing some of the LP constraints. Note that we can write

$$\begin{aligned} K(\emptyset) &\geq K(Y_1) \text{ by (monotonicity)} \\ K(Y_1) &\geq K(a(X), Y_1) + H(a(X)|Y_1) - \delta(\epsilon) \text{ by (monotonicity+), Fano's inequality, and } \delta(\epsilon) > 0, \\ K(a(X), Y_1) &\geq K(a(X), Y_2) \text{ by (monotonicity),} \\ K(a(X), Y_2) &\geq I(X; V, U_{a(X)Y_2}|a(X), Y_2) \text{ by (monotonicity+)} \text{ and } K(X, a(X), Y_2) = 0. \end{aligned}$$

Hence,  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  is lower bounded by

$$\inf_{V \in \mathcal{C}(X)} \inf_{U: \mathcal{C}(X) \rightarrow \mathcal{C}(X, V)} H(a(X)|Y_1) + I(X; V, U_{a(X)Y_2}|a(X), Y_2) - \delta(\epsilon).$$

By finding a cardinality constraint on  $(V, U_{a(X)Y_2})$  and letting  $\epsilon \rightarrow 0$ , we have the result.

Now we show that  $R_{ach}(\mathbf{D}) \leq R^{LN}(\mathbf{D})$ . By selecting the auxiliary random variables  $U_{\{1,2\}} = a(X)$ ,  $U_{\{1\}} = \emptyset$  and  $U_{\{2\}} = W$  in Lemma 1 and imposing the cardinality constraint  $|\mathcal{W}| \leq |\mathcal{X}| + 1$ , we have  $R_{ach}(\mathbf{D}) \leq R^{LN}(\mathbf{D})$ . ■

### C. Odd-cycle Index Coding Problem

The source  $\mathbf{X} = (X_1, \dots, X_m)$ , where  $m \geq 5$  is an odd number, is i.i.d. Bernoulli ( $\frac{1}{2}$ ) bits. The side information at decoder  $i$ ,  $i \in [m]$  is  $\mathbf{Y}_i = (X_{i-1}, X_{i+1})$ , where  $+$  and  $-$  in subscripts are modulo- $m$  operations<sup>10</sup>, and decoder  $i$  wishes to reconstruct  $X_i$  with a vanishing block error probability.

Although the achievability result Theorem 1 is for per-letter distortion constraints, it can be easily modified to accommodate block error probabilities. Let  $v \in \mathcal{V}$  be fixed. Then we select the messages  $\mathbf{U}_{\mathcal{J}_j}$ ,  $\mathcal{J}_j \in v$  such that

$$\mathbf{U}_{jk} = (X_j, X_k) \text{ for } j \in [m], k \equiv j + 1 \pmod{m} \quad (41)$$

and all of the other messages  $\mathbf{U}_{\mathcal{J}_j}$  are chosen to be  $\emptyset$ .<sup>11</sup> Let  $j \in [m]$ ,  $i \equiv j - 1 \pmod{m}$ ,  $k \equiv j + 1 \pmod{m}$ , and  $l \equiv k + 1 \pmod{m}$ . Then from the conditions in  $C_{ach}^{LP}$ , we can write

$$\begin{aligned} R_{jk} &\geq I(\mathbf{X}; \mathbf{U}_{jk}) - R'_{jk}, \text{ from condition 2) of } C_{ach}^{LP} \\ &= 2 - R'_{jk}, \\ R'_{jk} &\leq \min\{I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j), I(\mathbf{U}_{jk}; \mathbf{U}_{kl}, \mathbf{Y}_k)\}, \text{ from condition 3) of } C_{ach}^{LP} \\ &= 2, \\ R'_{ij} + R'_{jk} &\leq I(\mathbf{U}_{ij}; \mathbf{Y}_j) + I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j), \text{ from condition 3) of } C_{ach}^{LP} \\ &= 3. \end{aligned} \quad (42)$$

Then selecting  $R'_{jk} = \frac{3}{2}$  and  $R_{jk} = \frac{1}{2}$  satisfies the conditions of  $C_{ach}^{LP}$ . Hence, rate  $\frac{m}{2}$  is achievable. Also, in [10] it is shown that the LP lower bound in Theorem 7 gives  $\frac{m}{2}$  for the zero error case. From Theorem 8, we can conclude that the  $R_{lb}^I$  lower bound, which is a valid lower bound for vanishing error probability, also gives  $\frac{m}{2}$  which is the optimal rate for this problem.

Note that prior to this work, the minimax lower bound in Theorem 5,  $R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1})$ , was the state-of-the-art lower bound to  $R(\mathbf{D})$  for multiple decoders. The next lemma states that the minimax lower bound is strictly suboptimal for the odd-cycle index coding problem.

**Lemma 4.**  $\limsup_{\epsilon \rightarrow 0} R_{lb}^m(\mathbf{0} + \epsilon \mathbf{1}) < \frac{m}{2}$  for the odd-cycle index coding problem.

<sup>10</sup>Here  $x \pmod{m}$  is assumed to lie in  $[m]$ .

<sup>11</sup>We represent  $\mathbf{U}_{\{j,k\}}$  as  $\mathbf{U}_{jk}$  for ease of notation.

*Proof:* Firstly, note that  $\limsup_{\epsilon \rightarrow 0} R_{lb}^m(\mathbf{0} + \epsilon \mathbf{1}) \leq \frac{m}{2}$  since  $\frac{m}{2}$  is the optimal rate for the odd-cycle case. Also, notice that when we select  $(V, U_{Y_1}, \dots, U_{Y_m})$  such that  $V = \emptyset$ , and  $U_{Y_i} = X_i$ ,  $i \in [m]$ , the random variables  $(V, U_{Y_1}, \dots, U_{Y_m})$  are feasible in the optimization problem in Theorem 5 and  $\bar{R}_{lb}$  becomes

$$\max_{\sigma} [H(X_{\sigma(1)}|Y_{\sigma(1)}) + H(X_{\sigma(2)}|X_{\sigma(1)}, Y_{\sigma(1)}, Y_{\sigma(2)}) + \dots + H(X_{\sigma(m)}|X_{\sigma(1)}, \dots, X_{\sigma(m-1)}, Y)], \quad (43)$$

which is equal to the maximin lower bound for index coding in [11] implying  $\liminf_{\epsilon \rightarrow 0} R_{lb}^m(\mathbf{0} + \epsilon \mathbf{1}) = \bar{R}_{lb}$ . The fact that (43) must take a integer value concludes the proof. ■

Note that if we restrict the selection of auxiliary random variables to be a subset of the source  $\mathbf{X}$  in Theorem 2 of [12], the scheme in [12] becomes valid and each mutual information term in the optimization becomes an entropy of a subset of the source which gives an integer value. Hence, in this case the scheme in [12] gives an integer rate which is strictly suboptimal for this problem.

#### D. Odd-cycle Gaussian Rate-Distortion Problem

We finish with an instance that seems not to be solvable using existing lower bounds discussed in Section V. The problem setting we consider is analogous to the odd-cycle index coding problem [10], by taking each source component as an independent Gaussian random variable instead of uniform binary bits and considering a mean square error (MSE) distortion constraint on the reconstructions. Hence, we call it the *odd-cycle Gaussian* problem from now on. Specifically, the source  $\mathbf{X} = (X_1, \dots, X_m)$ , where  $m \geq 5$  is an odd number, is a Gaussian vector such that each component is independent of the others and has unit variance. The side information at decoder  $i$ ,  $i \in [m]$  is  $\mathbf{Y}_i = (X_{i-1}, X_{i+1})$ , where  $+$  and  $-$  in the subscripts are modulo- $m$  operations<sup>12</sup>, and decoder  $i$  wishes to reconstruct  $X_i$  subject to an MSE distortion constraint, i.e.,  $E[(X_i - \hat{X}_i)^2] \leq D$  for all  $i \in [m]$ .

**Theorem 12.** *The rate-distortion function,  $R^{IG}(\mathbf{D})$ , is*

$$R^{IG}(\mathbf{D}) = \frac{m}{4} \log \frac{1}{D}. \quad (44)$$

*Proof of Theorem 12: Achievability:* The achievability argument is obtained by using Theorem 2. Let  $v \in \mathcal{V}$  be fixed. We select the messages  $U_{\mathcal{J}_j}$  such that

$$\mathbf{U}_{jk} = (X_j + N_j, X_k + \bar{N}_k) \text{ for } j \in [m], k \equiv j + 1 \pmod{m} \quad (45)$$

and all the other messages  $U_{\mathcal{J}_j}$  are degenerate.<sup>13</sup> Here  $(N_i, \bar{N}_i)$ ,  $i \in [m]$  are Gaussian random variables with variance  $K_{N_i} = K_{\bar{N}_i} = \frac{2D}{1-D}$  and all  $N_i, \bar{N}_i$ 's are independent of each other and the source  $\mathbf{X}$ . All  $\mathbf{U}_{\mathcal{J}_j}$  satisfy conditions 1), 2) and 3) of  $C_{ach,v}^G(\mathbf{D})$  as well as condition 4) of  $C_{ach,v}^G(\mathbf{D})$  since  $K_{X_j|\mathbf{U}_{jk}, \mathbf{U}_{ij}, \mathbf{Y}_j} = (K_{X_j}^{-1} + K_{N_j}^{-1} + K_{\bar{N}_j}^{-1})^{-1} = D$ , where  $i = j - 1 \pmod{m}$ . Let  $j \in [m]$ ,  $i \equiv j - 1 \pmod{m}$ ,  $k \equiv j + 1 \pmod{m}$ , and  $l \equiv k + 1 \pmod{m}$ . Then from the conditions in  $C_{ach}^{LP}$ , we can write

$$R_{jk} \geq I(\mathbf{X}; \mathbf{U}_{jk}) - R'_{jk}, \text{ from condition 2) of } C_{ach}^{LP} \quad (46)$$

and since any disjoint sets of  $\mathbf{U}_{\mathcal{J}_j}$  are conditionally independent of each other given  $\mathbf{X}$ .

$$R'_{jk} \leq \min\{I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j), I(\mathbf{U}_{jk}; \mathbf{U}_{kl}, \mathbf{Y}_k)\}, \quad (47)$$

from condition 3) of  $C_{ach}^{LP}$ .

$$R'_{ij} + R'_{jk} \leq I(\mathbf{U}_{ij}; \mathbf{Y}_j) + I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j), \text{ by condition 3) of } C_{ach}^{LP}. \quad (48)$$

Note that the terms inside the minimum in (47) are equal to each other and also the encoding order of the messages does not affect the right-hand side of (48). Then using the chain rule, the mutual information terms in (46)–(48) can be written as

$$\begin{aligned} I(\mathbf{X}; \mathbf{U}_{jk}) &= I(X_j; X_j + N_j) + I(X_k; X_k + \bar{N}_k) \\ &= \log \frac{1+D}{2D}. \\ I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j) &= I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, X_i, X_k) \\ &= I(X_k + \bar{N}_k; X_k) + I(X_j + N_j; X_j + \bar{N}_j) \\ &= \frac{1}{2} \log \frac{1+D}{2D} + \frac{1}{2} \log \frac{(1+D)^2}{4D}. \\ I(\mathbf{U}_{ij}; \mathbf{Y}_j) + I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j) &= I(\mathbf{U}_{ij}; X_i, X_k) + I(\mathbf{U}_{jk}; \mathbf{U}_{ij}, \mathbf{Y}_j) \\ &= \frac{1}{2} \log \frac{1+D}{2D} + \frac{1}{2} \log \frac{1+D}{2D} + \frac{1}{2} \log \frac{(1+D)^2}{4D}. \end{aligned}$$

<sup>12</sup>Here as well,  $x \pmod{m}$  is assumed to lie in  $[m]$ .

<sup>13</sup>We represent  $U_{\{j,k\}}$  as  $U_{jk}$  for ease of notation.

Then selecting  $R'_{jk} = \frac{1}{2} \log \frac{1+D}{2D} + \frac{1}{4} \log \frac{(1+D)^2}{4D}$  and  $R_{jk} = \log \frac{1+D}{2D} - R'_{jk}$ ,  $j \in [m]$ ,  $k = j+1 \pmod m$  satisfies (46)–(48) and we take all other rates  $R_{\mathcal{S}_j}, R'_{\mathcal{S}_j}$  as 0. Hence, the achievable rate is

$$\begin{aligned} \sum_{i=1}^m R_{ij} &= m \left( \log \frac{1+D}{2D} - \frac{1}{2} \log \frac{1+D}{2D} - \frac{1}{4} \log \frac{(1+D)^2}{4D} \right) \\ &= \frac{m}{4} \left( 2 \log \frac{1+D}{2D} - \log \frac{(1+D)^2}{4D} \right) \\ &= \frac{m}{4} \log \frac{1}{D}. \end{aligned}$$

*Converse:* We utilize the computable relaxation of  $R_{lb}(\mathbf{D} + \epsilon \mathbf{1})$  in Theorem 4. Similar to the proof of [10, Theorem 5.1] we define the ordered sets:

$$\mathbf{O} = \{X_i : i \equiv 1 \pmod 2, i \neq m\}, \mathbf{O}^+ = \{X_i : i \leq m-2\}$$

$$\mathbf{E} = \{X_i : i \equiv 0 \pmod 2\}, \mathbf{E}^+ = \{X_i : 2 \leq i \leq m-1\}$$

$\mathbf{M} = \{X_i : 2 \leq i \leq m-2\}$ , and  $\mathbf{S} = \mathbf{X} \setminus (\mathbf{M} \cup X_m)$ . Note that  $(\mathbf{O}^+ \setminus \mathbf{O}) \cap (\mathbf{E}^+ \setminus \mathbf{E}) = \emptyset$  and  $\mathbf{M} = (\mathbf{O}^+ \setminus \mathbf{O}) \cup (\mathbf{E}^+ \setminus \mathbf{E})$ . Also, define  $R(D) = \frac{1}{2} \log \frac{1}{D}$ . Then using the conditions of the LP in Theorem 4 we can obtain the following inequalities

$$K(\emptyset) \geq K(\mathbf{O}) \quad \text{by (monotonicity)} \quad (49)$$

$$K(\emptyset) \geq K(\mathbf{E}) \quad \text{by (monotonicity)} \quad (50)$$

$$K(\emptyset) \geq K(X_m) \quad \text{by (monotonicity)} \quad (51)$$

$$K(\mathbf{O}) \geq K(\mathbf{O}^+) + \sum_{X_i \in \mathbf{O}^+ \setminus \mathbf{O}} R(D + \epsilon) \quad (52)$$

$$K(\mathbf{E}) \geq K(\mathbf{E}^+) + \sum_{X_i \in \mathbf{E}^+ \setminus \mathbf{E}} R(D + \epsilon) \quad (53)$$

$$K(\mathbf{O}^+) + K(\mathbf{E}^+) \geq K(\mathbf{M}) + K(\mathbf{X}) + R(D + \epsilon) \quad (54)$$

$$K(\mathbf{M}) + K(X_m) \geq K(\emptyset) + K(\mathbf{X}) + \sum_{X_i \in \mathbf{S}} R(D + \epsilon) \quad (55)$$

where (52) is due to the following. By *monotonicity*<sub>+</sub>, we have  $K(\mathbf{O}) \geq K(\mathbf{O}^+) + R(D_{\mathbf{O}} + \epsilon)$ . We can see that  $\sum_{X_i \in \mathbf{O}^+ \setminus \mathbf{O}} R(D + \epsilon)$  is an upper bound to  $R(D_{\mathbf{O}} + \epsilon)$  by selecting the auxiliary random variable  $V = \{X_i + N_i | i \in [m]\}$ , where  $N_i$  is independent of  $\mathbf{X}$  and all  $N_j$ 's,  $j \neq i$ , in the minimization in Notation 8. Also, utilizing the chain rule one can verify that  $R(D_{\mathbf{O}} + \epsilon) \geq \sum_{X_i \in \mathbf{O}^+ \setminus \mathbf{O}} R(D + \epsilon)$ . By following a similar procedure to that used to obtain (52), we can also obtain (53). Furthermore, (54) and (55) are due to *submodularity* and *monotonicity*<sub>+</sub>. If we add inequalities (49)–(55) side-by-side, we obtain  $2K(\emptyset) \geq mR(D + \epsilon)$ . Taking  $\epsilon \rightarrow 0$  gives the result. ■

Recall that prior to the LP lower bound introduced here, the lower bound in Theorem 5,  $R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1})$ , was the state-of-the-art lower bound to  $R(\mathbf{D})$ . The next lemma shows that  $R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1})$  gives  $\frac{m-1}{4} \log(\frac{1}{D})$  for the odd-cycle Gaussian problem and is thus not tight.

**Lemma 5.** *For the odd-cycle Gaussian problem,  $\liminf_{\epsilon \rightarrow 0} R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) = \frac{m-1}{4} \log(\frac{1}{D})$ .*

*Proof of Lemma 5:* The proof is given in the Appendix B. ■

#### ACKNOWLEDGMENT

The authors wish to thank Robert Kleinberg for several helpful comments. This work was supported by the US National Science Foundation under grant CCF-1617673.

#### APPENDIX A

*Proof of Theorem 1:* Let  $\epsilon > 0$ ,  $v \in \mathcal{V}$  be given and suppose the joint distribution of  $(\mathcal{U}, X, Y_1, \dots, Y_m)$  in  $C_{ach,v}(\mathbf{D})$ , denoted by  $p$ , is fixed. The scheme consists of three main steps; namely, code construction, encoding and decoding. First we explain each step then show that the resulting rate is  $\mathbf{D}$ -achievable.

Code construction and encoding are similar to the proof of the achievable scheme in [12], which depends on  $\epsilon$ -letter typicality [17] arguments. Here we use the lowercase letter  $z$  to denote a realization of a random variable  $Z$ .

*Code Construction :* A codebook, denoted by  $\mathcal{C}_{\mathcal{S}_j}$ , of size  $2^{n(R_{\mathcal{S}_j} + R'_{\mathcal{S}_j})}$  is created for each set  $\mathcal{S}_j \in v$  in the following way. Let  $\mathbf{k}_{\mathcal{S}_j} = (k_{\mathcal{S}_j}, k'_{\mathcal{S}_j})$ , where  $k_{\mathcal{S}_j} \in [2^{nR_{\mathcal{S}_j}}]$  and  $k'_{\mathcal{S}_j} \in [2^{nR'_{\mathcal{S}_j}}]$ . A codeword  $u_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j}) \in \mathcal{U}_{\mathcal{S}_j}^n$  of length  $n$  is created by drawing each component from  $\mathcal{U}_{\mathcal{S}_j}$  with respect to  $p(u_{\mathcal{S}_j})$  in an i.i.d. way.



*Encoding* : Let  $0 < \epsilon_0 < \dots < \epsilon_{2^m+1}$  be sufficiently small and  $x^n \in \mathcal{X}^n$  be given to the encoder. Then encoding is performed in  $2^m - 1$  stages. Specifically, at stage  $j$  encoder picks  $\mathcal{C}^{\mathcal{S}_j}$  and searches for an index  $\mathbf{k}_{\mathcal{S}_j}$  such that  $u_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j})$  is  $\epsilon_j$ -letter typical with  $x^n$  and

$$u_{\mathcal{S}_j}^- = \{u_{\mathcal{S}_i}(\mathbf{k}_{\mathcal{S}_i}) | i < j\}. \quad (56)$$

If such a  $\mathbf{k}_{\mathcal{S}_j}$  (or multiple such  $\mathbf{k}_{\mathcal{S}_j}$ ) exists then the encoder picks one of them arbitrarily and sends the bin index  $k_{\mathcal{S}_j}$  to the decoders. Otherwise the encoder picks a codeword randomly and sends the corresponding bin index.

*Decoding* : We apply simultaneous decoding [1, Section 4]. Consider decoder  $l$ . It forms reconstructions of all its messages,  $u_{\mathcal{D}_l}(\hat{\mathbf{k}}_{\mathcal{D}_l}) = \{u_{\mathcal{S}_j}(\hat{\mathbf{k}}_{\mathcal{S}_j}) | \mathcal{S}_j \in \mathcal{D}_l\}$ , where  $\hat{\mathbf{k}}_{\mathcal{D}_l} = \{\hat{\mathbf{k}}_{\mathcal{S}_j} | \mathcal{S}_j \in \mathcal{D}_l\}$ <sup>14</sup>, in the following way. Decoder  $l$  takes the set of bin indices  $k_{\mathcal{D}_l} = \{k_{\mathcal{S}_j} | \mathcal{S}_j \in \mathcal{D}_l\}$  then looks for a set of indices  $\tilde{\mathbf{k}}_{\mathcal{D}_l}$  such that

$$\tilde{k}_{\mathcal{S}_j} = k_{\mathcal{S}_j} \text{ for all } \mathcal{S}_j \in \mathcal{D}_l \text{ and} \quad (57)$$

$$u_{\mathcal{D}_l}(\tilde{\mathbf{k}}_{\mathcal{D}_l}) \text{ are } \epsilon_{l^*+1}\text{-letter typical with } y_l^n, \quad (58)$$

where  $l^* = \max_{j: \mathcal{S}_j \in \mathcal{D}_l} j$ . Note that if no error occurs at the encoder,  $u_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l})$  is  $\epsilon_{l^*}$ -typical with  $x^n$ . If there is more than one set of codewords  $u_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j})$ ,  $\mathcal{S}_j \in \mathcal{D}_l$  whose indices,  $\tilde{\mathbf{k}}_{\mathcal{S}_j}$ , satisfy (57) and (58), decoder  $l$  selects one arbitrarily and sets  $\hat{\mathbf{k}}_{\mathcal{S}_j} = \tilde{\mathbf{k}}_{\mathcal{S}_j}$ . If decoder  $l$  cannot find any such set of indices, it sets  $\hat{\mathbf{k}}_{\mathcal{D}_l}$  to  $\mathbf{1}$  (i.e., it declares an error). Since the joint distribution of  $(\mathcal{U}, X, Y_1, \dots, Y_m)$  is in  $C_{ach,v}(\mathbf{D})$ , we can find a function  $g_l(\cdot, \cdot)$  such that  $g_l(u_{\mathcal{D}_l i}(\hat{\mathbf{k}}_{\mathcal{D}_l}), y_{li}) = \hat{x}_{li}$ , where  $u_{\mathcal{D}_l i}(\hat{\mathbf{k}}_{\mathcal{D}_l})$ ,  $y_{li}$  and  $\hat{x}_{li}$  are the  $i^{\text{th}}$  components of  $u_{\mathcal{D}_l}(\hat{\mathbf{k}}_{\mathcal{D}_l})$ ,  $y_l^n$  and  $\hat{x}_l^n$ , respectively.

Now we analyze the error probabilities at the encoding and decoding steps, respectively.

*Error Analysis for Encoder* : Note that encoding process is correct if the following is satisfied:

1. At each encoding stage  $j$ , we can find  $U_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j})$  such that it is  $\epsilon_j$ -jointly typical with  $(U_{\mathcal{S}_j}^-, X^n)$  i.e.,

$$C_{\mathcal{S}_j} = \left\{ \exists \mathbf{k}_{\mathcal{S}_j} \text{ such that } u_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j}) \in \mathcal{T}_{\epsilon_j}^{(n)}(p | U_{\mathcal{S}_j}^-, X^n) \right\}. \quad (59)$$

Then the probability of error at the encoder,  $\Pr(E)$  can be expressed as

$$\begin{aligned} \Pr(E) &= \Pr((C_{\mathcal{S}_1} \cap \dots \cap C_{\mathcal{S}_{2^m-1}})^c) \\ &= \Pr(C_{\mathcal{S}_1}^c \cup \dots \cup C_{\mathcal{S}_{2^m-1}}^c) \\ &= \Pr((C_{\mathcal{S}_1}^c \cap \bar{C}^1) \cup \dots \cup (C_{\mathcal{S}_{2^m-1}}^c \cap \bar{C}^{2^m-1})), \end{aligned} \quad (60)$$

where  $\bar{C}^j$  is defined as  $\bigcap_{i=1}^{j-1} C_{\mathcal{S}_i}$  for all  $j \in [2^m - 1] \setminus \{1\}$  and  $\bar{C}^1 = \emptyset$ . Then from (60) and the union bound, we can write

$$\begin{aligned} \Pr(E) &\leq \Pr(C_{\mathcal{S}_1}^c \cap \bar{C}^1) + \dots + \Pr(C_{\mathcal{S}_{2^m-1}}^c \cap \bar{C}^{2^m-1}) \\ &\leq \Pr(C_{\mathcal{S}_1}^c | \bar{C}^1) + \dots + \Pr(C_{\mathcal{S}_{2^m-1}}^c | \bar{C}^{2^m-1}). \end{aligned} \quad (61)$$

Note that  $\Pr(C_{\mathcal{S}_j}^c | \bar{C}^j)$ ,  $j \in [2^m - 1]$  represents the probability of the event that there is no  $U_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j})$   $\epsilon_j$ -jointly typical with  $(U_{\mathcal{S}_j}^-, X^n)$  given that for each  $i < j$  we find  $U_{\mathcal{S}_i}(\mathbf{k}_{\mathcal{S}_i})$  such that  $U_{\mathcal{S}_i}(\mathbf{k}_{\mathcal{S}_i})$  is  $\epsilon_i$ -jointly typical with  $(U_{\mathcal{S}_i}^-, X^n)$ , i.e.,

$$\Pr(C_{\mathcal{S}_j}^c | \bar{C}^j) = \Pr\left(\forall \mathbf{k}_{\mathcal{S}_j}, U_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j}) \notin \mathcal{T}_{\epsilon_j}^{(n)}(p | U_{\mathcal{S}_j}^-, X^n) | (U_{\mathcal{S}_j}^-, X^n) \in \mathcal{T}_{\epsilon_{j-1}}^{(n)}(p)\right).$$

From Lemma 8 in Appendix C and the inequality  $(1 - \alpha)^\beta < e^{-\alpha\beta}$ , we can write

$$\begin{aligned} \Pr(C_{\mathcal{S}_j}^c | \bar{C}^j) &< e^{-\left[ (1 - \delta_{\epsilon_{j-1}, \epsilon_j}(n)) 2^{-n(I(X, U_{\mathcal{S}_j}^-; U_{\mathcal{S}_j}) + 2\epsilon_j H(U_{\mathcal{S}_j}))} \right] 2^{n(R_{\mathcal{S}_j} + R'_{\mathcal{S}_j})}} \\ &= e^{-\left[ (1 - \delta_{\epsilon_{j-1}, \epsilon_j}(n)) 2^{n((R_{\mathcal{S}_j} + R'_{\mathcal{S}_j}) - I(X, U_{\mathcal{S}_j}^-; U_{\mathcal{S}_j}) - 2\epsilon_j H(U_{\mathcal{S}_j}))} \right]}, \end{aligned} \quad (62)$$

where  $\delta_{\epsilon_{j-1}, \epsilon_j}(n) \rightarrow 0$  as  $n \rightarrow \infty$ . Note that when  $H(U_{\mathcal{S}_j}) = 0$ ,  $\Pr(C_{\mathcal{S}_j}^c | \bar{C}^j)$  is equal to zero. Then  $\Pr(C_{\mathcal{S}_j}^c | \bar{C}^j) < \frac{\epsilon'}{2^m}$  if  $n \geq n_1(\epsilon', \epsilon_j H(U_{\mathcal{S}_j}))$ , and

$$R_{\mathcal{S}_j} + R'_{\mathcal{S}_j} \geq I(X, U_{\mathcal{S}_j}^-; U_{\mathcal{S}_j}) + 3\epsilon_j H(U_{\mathcal{S}_j}). \quad (63)$$

Hence, if  $(R_{\mathcal{S}_j}, R'_{\mathcal{S}_j})$  satisfy the condition in (63) for all  $j \in [2^m - 1]$ , from (61) we can conclude that the probability of error at the encoder satisfies

$$\Pr(E) < \frac{2^m - 1}{2^m} \epsilon' \quad (64)$$

<sup>14</sup> Since  $v \in \mathcal{V}$  is an ordered list, it induces an order on sets  $\mathcal{S}_j$ . Hence we can take  $\hat{\mathbf{k}}_{\mathcal{D}_l}$  as an ordered set and assume an ordered set structure.

when  $n \geq N_1$  where  $N_1 = \max_{j \in [2^m-1]} n_1(\epsilon', \epsilon_j H(U_{\mathcal{S}_j}))$ .

*Error Analysis for Decoders :* Let us focus on decoder  $l$  for some fixed  $l \in [m]$ . Decoding at this decoder is successful if the following conditions are satisfied:

1. There is no error at the encoder.
2. The source and the side information are  $\epsilon_0$ -typical, i.e.,

$$D_0 = \left\{ (X^n, Y_1^n, \dots, Y_m^n) \in \mathcal{T}_{\epsilon_0}^{(n)}(p) \right\}. \quad (65)$$

3. The set of codewords  $U_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l}) = \{U_{\mathcal{S}_j}(\mathbf{k}_{\mathcal{S}_j}) | \mathcal{S}_j \in \mathcal{D}_l\}$  chosen by the encoder are  $\epsilon_{l^*+1}$ -letter typical with  $Y_l^n$ , i.e.,

$$D_{1,l} = \left\{ (U_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l}), X^n, Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p) \right\}. \quad (66)$$

4. Within the received bins  $k_{\mathcal{D}_l} = \{k_{\mathcal{S}_j} | \mathcal{S}_j \in \mathcal{D}_l\}$ , decoder  $l$  can find a unique set of codewords,  $U_{\mathcal{D}_l}(\hat{\mathbf{k}}_{\mathcal{D}_l}) = \{U_{\mathcal{S}_j}(\hat{\mathbf{k}}_{\mathcal{S}_j}) | \hat{k}_{\mathcal{S}_j} = k_{\mathcal{S}_j}, \mathcal{S}_j \in \mathcal{D}_l\}$ , such that  $U_{\mathcal{D}_l}(\hat{\mathbf{k}}_{\mathcal{D}_l})$  are  $\epsilon_{l^*+1}$ -letter typical with  $Y_l^n$ , i.e.,

$$D_{2,l} = \left\{ \exists \tilde{\mathbf{k}}_{\mathcal{D}_l} \neq \mathbf{k}_{\mathcal{D}_l} \text{ such that } \tilde{k}_{\mathcal{D}_l} = k_{\mathcal{D}_l}, (U_{\mathcal{D}_l}(\tilde{\mathbf{k}}_{\mathcal{D}_l}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p) \right\}. \quad (67)$$

Then we can write the probability of error at decoder  $l$ , denoted by  $\Pr(D_{err,l})$ , as

$$\begin{aligned} \Pr(D_{err,l}) &= \Pr((E^c \cap D_0 \cap D_{1,l} \cap D_{2,l})^c) \\ &= \Pr(E \cup D_0^c \cup D_{1,l}^c \cup D_{2,l}^c) \\ &= \Pr(\bar{E} \cup (D_{1,l}^c \cap \bar{E}^c) \cup (D_{2,l}^c \cap \bar{E}^c \cap D_{1,l})), \text{ where } \bar{E} = E \cup D_0^c, \\ &\leq \Pr(\bar{E}) + \Pr(D_{1,l}^c \cap \bar{E}^c) + \Pr(D_{2,l}^c \cap \bar{E}^c \cap D_{1,l}). \end{aligned} \quad (68)$$

First we analyze  $\Pr(\bar{E})$ . By Lemma 7 in Appendix C,  $\Pr(D_0^c) < \delta_{\epsilon_0}(n)$  where  $\delta_{\epsilon_0}(n) \rightarrow 0$  as  $n \rightarrow \infty$ . Then we can find  $n_2(\epsilon', \delta_{\epsilon_0})$ ,  $\epsilon' > 0$  such that if  $n \geq n_2(\epsilon', \delta_{\epsilon_0})$ ,  $\Pr(D_0^c) < \frac{\epsilon'}{2^m}$ . Hence, from (64) and the union bound, we have

$$\Pr(\bar{E}) \leq \Pr(E) + \Pr(D_0^c) < \epsilon' \quad (69)$$

when  $n \geq \max\{n_2(\epsilon', \delta_{\epsilon_0}), N_1\}$ .

Now we focus on  $\Pr(D_{1,l}^c \cap \bar{E}^c)$  and  $\Pr(D_{2,l}^c \cap \bar{E}^c \cap D_{1,l})$ . We can upper bound  $\Pr(D_{1,l}^c \cap \bar{E}^c)$  by

$$\Pr\left((U_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l}), X^n, Y_l^n) \notin \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p) | (U_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l}), X^n) \in \mathcal{T}_{\epsilon_{l^*}}^{(n)}(p)\right). \quad (70)$$

By Lemma 9 in Appendix C, the probability in (70) is less than or equal to  $\delta_{\epsilon_{l^*} \epsilon_{l^*+1}}(n)$  which goes to 0 as  $n \rightarrow \infty$ . Hence,  $\Pr(D_{1,l}^c \cap \bar{E}^c) < \epsilon'$  if  $n \geq n_3(\epsilon', \delta_{\epsilon_{l^*} \epsilon_{l^*+1}})$ .

Now we consider  $\Pr(D_{2,l}^c \cap \bar{E}^c \cap D_{1,l})$ . Note that event  $D_{2,l}^c$  can be rewritten as

$$\begin{aligned} D_{2,l}^c &= \bigcup_{\mathcal{D}'_l: \mathcal{D}'_l \subseteq \mathcal{D}_l, \mathcal{D}'_l \neq \emptyset} F_{\mathcal{D}'_l}, \text{ where} \\ F_{\mathcal{D}'_l} &= \left\{ \exists \tilde{\mathbf{k}}_{\mathcal{D}_l} \text{ such that } \tilde{\mathbf{k}}_{\mathcal{S}_j} \neq \mathbf{k}_{\mathcal{S}_j} \text{ for all } \mathcal{S}_j \in \mathcal{D}'_l, \tilde{k}_{\mathcal{D}'_l} = k_{\mathcal{D}'_l}, \tilde{\mathbf{k}}_{\mathcal{S}_j} = \mathbf{k}_{\mathcal{S}_j} \text{ for all } \mathcal{S}_j \in \mathcal{D}_l \setminus \mathcal{D}'_l \text{ and} \right. \\ &\quad \left. (U_{\mathcal{D}_l}(\tilde{\mathbf{k}}_{\mathcal{D}_l}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p) \right\}. \end{aligned}$$

Using the union bound, we can write

$$\Pr(D_{2,l}^c \cap \bar{E}^c \cap D_{1,l}) \leq \sum_{\mathcal{D}'_l: \mathcal{D}'_l \subseteq \mathcal{D}_l, \mathcal{D}'_l \neq \emptyset} \Pr(F_{\mathcal{D}'_l} \cap \bar{E}^c \cap D_{1,l}). \quad (71)$$

Notice that  $F_{\mathcal{D}'_l} \cap \bar{E}^c \cap D_{1,l}$  denotes the error event that there is no error at the encoder and the source and side information are  $\epsilon_0$ -typical (event  $\bar{E}^c$ ), and decoder  $l$  can find a set of indices  $\{\hat{\mathbf{k}}_{\mathcal{S}_j} | \mathcal{S}_j \in \mathcal{D}_l\}$  such that  $U_{\mathcal{D}_l}(\hat{\mathbf{k}}_{\mathcal{D}_l})$  are  $\epsilon_{l^*+1}$ -jointly typical with  $(X^n, Y_l^n)$  (event  $D_{1,l}$ ); however the particular subset  $\mathbf{k}_{\mathcal{D}'_l} = \{\mathbf{k}_{\mathcal{S}_j} | \mathcal{S}_j \in \mathcal{D}'_l\}$  of those indices is not unique (event  $F_{\mathcal{D}'_l}$ ). Now we bound each term inside the summation in (71). To do this, first we define an event  $\bar{F}_{\mathcal{D}'_l}$  by replacing the typical set  $\mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p)$  in event  $F_{\mathcal{D}'_l}$  with  $\mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p)$ . In other words,

$$\begin{aligned} \bar{F}_{\mathcal{D}'_l} &= \left\{ \exists \tilde{\mathbf{k}}_{\mathcal{D}_l} \text{ such that } \tilde{\mathbf{k}}_{\mathcal{S}_j} \neq \mathbf{k}_{\mathcal{S}_j} \text{ for all } \mathcal{S}_j \in \mathcal{D}'_l, \tilde{k}_{\mathcal{D}'_l} = k_{\mathcal{D}'_l}, \tilde{\mathbf{k}}_{\mathcal{S}_j} = \mathbf{k}_{\mathcal{S}_j} \text{ for all } \mathcal{S}_j \in \mathcal{D}_l \setminus \mathcal{D}'_l \text{ and} \right. \\ &\quad \left. (U_{\mathcal{D}_l}(\tilde{\mathbf{k}}_{\mathcal{D}_l}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p) \right\}, \end{aligned}$$

giving  $F_{\mathcal{D}'_l} \subseteq \bar{F}_{\mathcal{D}'_l}$ . Let  $S_1 = \{\tilde{\mathbf{k}}_{\mathcal{D}'_l} | \tilde{k}'_{\mathcal{S}_j} \neq k'_{\mathcal{S}_j}, \tilde{k}_{\mathcal{D}'_l} = k_{\mathcal{D}'_l}, \forall \mathcal{S}_j \in \mathcal{D}'_l\}$  and  $S_2 = \{\tilde{\mathbf{k}}_{\mathcal{D}'_l} | \tilde{k}_{\mathcal{S}_j} = 1, \forall \mathcal{S}_j \in \mathcal{D}'_l\}$ . Then we can write

$$\Pr(\bar{F}_{\mathcal{D}'_l} \cap \bar{E}^c \cap D_{1,l}) \leq \Pr\left(\bigcup_{S_1} U_{\mathcal{D}'_l}(\tilde{\mathbf{k}}_{\mathcal{D}'_l}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}'_l}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}'_l}), Y_l^n) | (U_{\mathcal{D}_l \setminus \mathcal{D}'_l}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}'_l}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p)\right) \quad (72)$$

$$\leq \Pr\left(\bigcup_{S_2} U_{\mathcal{D}'_l}(\tilde{\mathbf{k}}_{\mathcal{D}'_l}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}'_l}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}'_l}), Y_l^n) | (U_{\mathcal{D}_l \setminus \mathcal{D}'_l}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}'_l}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p)\right), \quad (73)$$

where (73) is obtained by using Lemma 10 in Appendix C. Then due to the union bound of probabilities we can write

$$\begin{aligned} \Pr(\bar{F}_{\mathcal{D}'_l} \cap \bar{E}^c \cap D_{1,l}^c) &\leq \sum_{S_2} \Pr\left(U_{\mathcal{D}'_l}(\tilde{\mathbf{k}}_{\mathcal{D}'_l}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}'_l}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}'_l}), Y_l^n) | (U_{\mathcal{D}_l \setminus \mathcal{D}'_l}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}'_l}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p)\right) \\ &\leq 2^{n \sum_{\mathcal{S}_j \in \mathcal{D}'_l} R'_{\mathcal{S}_j}} 2^{-n(\sum_{\mathcal{S}_j \in \mathcal{D}'_l} H(U_{\mathcal{S}_j}) - H(U_{\mathcal{D}'_l} | U_{\mathcal{D}_l \setminus \mathcal{D}'_l}, Y_l) - 2\epsilon_{l^*+2}(\sum_{\mathcal{S}_j \in \mathcal{D}'_l} H(U_{\mathcal{S}_j})))}, \text{ from Corollary 1.} \end{aligned} \quad (74)$$

Note that  $R'_{\mathcal{S}_j} \geq 0$ , for all  $j \in [2^m - 1]$  and when each  $R'_{\mathcal{S}_j} = 0$ ,  $\mathcal{S}_j \in \mathcal{D}'_l$ , there is only one codeword  $U(\mathbf{k}_{\mathcal{S}_j})$ ,  $\mathcal{S}_j \in \mathcal{D}'_l$  in each bin. Then, from (72)  $\Pr(\bar{F}_{\mathcal{D}'_l} \cap \bar{E}^c \cap D_{1,l}) = 0$  in this case. Also, when each  $H(U_{\mathcal{S}_j}) = 0$ ,  $\mathcal{S}_j \in \mathcal{D}'_l$ ,  $\Pr(\bar{F}_{\mathcal{D}'_l} \cap \bar{E}^c \cap D_{1,l})$  is equal to 0.

Thus from (74), if

$$\sum_{\mathcal{S}_j \in \mathcal{D}'_l} R'_{\mathcal{S}_j} \leq \max\left\{\left(\sum_{\mathcal{S}_j \in \mathcal{D}'_l} H(U_{\mathcal{S}_j})\right) - H(U_{\mathcal{D}'_l} | U_{\mathcal{D}_l \setminus \mathcal{D}'_l}, Y_l) - 3\epsilon_{l^*+2} \left(\sum_{\mathcal{S}_j \in \mathcal{D}'_l} H(U_{\mathcal{S}_j})\right), 0\right\} \quad (75)$$

and  $n \geq n_4(\epsilon', \epsilon_{l^*+2}, H(U_{\mathcal{S}_j}))$ ,  $\Pr(D_{2,l}^c \cap \bar{E}^c \cap D_{1,l}^c) < \frac{\epsilon'}{2^{|\mathcal{D}'_l|}}$ . Then from (68), if  $R'_{\mathcal{S}_j}$  satisfies (75) for all  $\mathcal{D}_l$ ,  $l \in [m]$  and  $n > N$ , where  $N = \max\{N_1, n_2(\epsilon', \delta_{\epsilon_0}), n_3(\epsilon', \delta_{\epsilon_{l^*+1}}), \max_{l \in [m]} \{n_4(\epsilon', \epsilon_{l^*+2}, H(U_{\mathcal{S}_j}))\}\}$

$$\Pr(D_{err,l}) < 3\epsilon'. \quad (76)$$

Let

$$D_{err} = \cup_{l \in [m]} D_{err,l}$$

denote the event that there is a decoding error at some decoder. By (76) and the union bound we have

$$\Pr(D_{err}) < 3\epsilon' m. \quad (77)$$

Thus there must exist a single code in the ensemble for which (77) holds. Now we focus on the distortion constraints at the decoder for this particular code. Assuming that there is no error occurring at the encoder and the decoders (corresponding to event  $E^c \cap D_{err,l}^c$ ), decoder  $l$  can find a unique  $u_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l})$  such that  $(u_{\mathcal{D}_l}(\mathbf{k}_{\mathcal{D}_l}), y_l^n, x^n)$  is  $\epsilon_{l^*+1}$ -jointly typical and it can reconstruct  $\hat{x}_l^n$  symbol-by-symbol through  $\hat{x}_{li} = g_l(u_{\mathcal{D}_l i}, y_{li})$ ,  $i \in [n]$ . Then using the arguments in [17, page 57] we can bound the average distortion at decoder  $l$  by

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n d_l(x_i, \hat{x}_{li}) &= \sum_{i=1}^n d_l(x_i, g_l(u_{\mathcal{D}_l i}, y_{li})) \\ &\leq E[d_l(X, g_l(U_{\mathcal{D}_l}, Y_l))] + \epsilon_{l^*+1} D_{l,max} \\ &\leq D_l + \epsilon_{l^*+1} D_{l,max}, \end{aligned} \quad (78)$$

where  $D_{l,max}$  is the maximum distortion that  $d_l(\cdot, \cdot)$  can give. Then the expected distortion at decoder  $l$  can be bounded by

$$\begin{aligned} E\left[\frac{1}{n} \sum_{i=1}^n d_l(x_i, \hat{x}_{li})\right] &\leq (D_l + \epsilon_{l^*+1} D_{l,max}) \Pr(E^c \cap D_{err,l}^c) + D_{l,max} \Pr(E \cup D_{err,l}) \\ &\leq D_l + D_{l,max}(\epsilon_{l^*+1} + \Pr(E \cup D_{err,l})) \\ &< D_l + D_{l,max}(\epsilon_{l^*+1} + 4\epsilon' m), \end{aligned} \quad (79)$$

where (79) holds if  $n > N$  and  $(R_{\mathcal{S}_j}, R'_{\mathcal{S}_j})$ ,  $\mathcal{S}_j \subseteq [m]$  satisfy the conditions in (63), (75), and the following non-negativity conditions :

$$R_{\mathcal{S}_j} \geq 0, \text{ for all } j \in [2^m - 1] \quad (80)$$

$$R'_{\mathcal{S}_j} \geq 0, \text{ for all } j \in [2^m - 1]. \quad (81)$$

Thus for all sufficiently large  $n$ , there exists a code whose expected distortion at decoder  $l$  satisfies (79) and whose rate does not exceed

$$\inf \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j} \quad (82)$$

subject to  $:R_{\mathcal{S}_j}, R'_{\mathcal{S}_j}, j \in [2^m - 1]$  satisfying (63), (75), (80), and (81).

**Lemma 6.** Let  $0 < \epsilon_0 < \epsilon_1 < \dots < \epsilon_{2^m+1}$ , and  $U_{\mathcal{S}_j}, \mathcal{S}_j \in v$ , be as in the proof of Theorem 1. For  $\gamma \geq 0$ , consider the following linear program:

$$\tilde{R}(\gamma) = \inf_{C_{ach}^{LP}(\gamma)} \sum_{j=1}^{2^m-1} R_{\mathcal{S}_j}, \quad (83)$$

where  $C_{ach}^{LP}(\gamma)$  denotes the set of  $R_{\mathcal{S}_j}$  and  $R'_{\mathcal{S}_j}$  such that

- 1)  $R_{\mathcal{S}_j} \geq 0$  and  $R'_{\mathcal{S}_j} \geq 0$ , for all  $j \in [2^m - 1]$ ;
- 2)  $R_{\mathcal{S}_j} + R'_{\mathcal{S}_j} \geq I(X, U_{\mathcal{S}_j}^-; U_{\mathcal{S}_j}) + 3\gamma$ , for all  $j \in [2^m - 1]$ ;
- 3) For each decoder  $l$ ,  $l \in [m]$

$$\sum_{\mathcal{S}_j \in \mathcal{D}_l'} R'_{\mathcal{S}_j} \leq \max \left\{ \left( \sum_{\mathcal{S}_j \in \mathcal{D}_l'} H(U_{\mathcal{S}_j}) \right) - H(U_{\mathcal{D}_l'} | U_{\mathcal{D}_l \setminus \mathcal{D}_l'}, Y_l) - 3(2^m - 1)\gamma, 0 \right\}.$$

Then  $\tilde{R}(\gamma)$  is continuous at  $\gamma = 0$  and is greater than or equal to the optimal value in (82) if

$$\gamma \geq \epsilon_{2^m+1} \max_{U_{\mathcal{S}_j}} H(U_{\mathcal{S}_j}). \quad (84)$$

*Proof of Lemma 6:* Note that when  $\gamma = 0$ ,  $C_{ach}^{LP}(\gamma)$  is equal to  $C_{ach}^{LP}$ . Also, since the alphabets are finite,  $C_{ach}^{LP}(\gamma)$  is nonempty for any  $\gamma \geq 0$ . The continuity of  $\tilde{R}(\gamma)$  in  $\gamma$  then follows from standard results on the continuity of LPs [13]. The relation with (82) follows by noting that  $C_{ach}^{LP}(\gamma)$  is contained in the set defined by the constraints (63), (75), (80), and (81), whenever (84) holds. ■

Now given  $\epsilon > 0$ , choose  $0 < \epsilon_0 < \epsilon_1 < \dots < \epsilon_{2^m+1}$ ,  $\epsilon'$  and  $\gamma$  such that

$$\begin{aligned} D_{l,max}(\epsilon_{l^*+1} + 4\epsilon'm) &< \epsilon \text{ for all } l \in [m] \\ \gamma &\geq \epsilon_{2^m+1} \max_{U_{\mathcal{S}_j}} H(U_{\mathcal{S}_j}) \end{aligned}$$

and  $\tilde{R}(\gamma) < \tilde{R}(0) + \epsilon$ . Then we have that for all sufficiently large  $n$ , there exists a code with rate at most  $\tilde{R}(0) + \epsilon$  whose expected distortion at decoder  $l$  is at most  $D_l + \epsilon$ . It follows that  $\tilde{R}(0)$  is  $\mathbf{D}$ -achievable as desired and hence  $R'_{ach}(\mathbf{D})$  is  $\mathbf{D}$ -achievable. Lastly, since  $R(\mathbf{D})$  is a convex function with respect to  $\mathbf{D}$  and it is upper bounded by  $R'_{ach}(\mathbf{D})$ ,  $R(\mathbf{D})$  must lie beneath the lower convex envelope of  $R'_{ach}(\mathbf{D})$ . ■

## APPENDIX B

*Proof of Lemma 5:* Let  $\epsilon > 0$  be given. Note that since each side information variable is a function of the source, the set  $\bar{P}$  in Theorem 5 contains only one element. Then let us select  $V = \emptyset$  and  $U_{Y_i} = X_i + N_i$  for all  $i \in [m]$  where  $N_i$  is independent of  $\mathbf{X}$  and the rest of the  $N_j$ 's,  $j \neq i$  and is such that  $K_{X_i|U_{Y_i}} = D + \epsilon$ . Then the  $U_{Y_i}$ 's are feasible in the optimization in Theorem 5 and we can write

$$R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) \leq \max_{\sigma} R_{\sigma} \quad (85)$$

where

$$\begin{aligned} R_{\sigma} &= I(\mathbf{X}; U_{Y_{\sigma(1)}} | \mathbf{Y}_{\sigma(1)}) + I(\mathbf{X}; U_{Y_{\sigma(2)}} | U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots \\ &\quad + I(\mathbf{X}; U_{Y_{\sigma(m)}} | U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(m-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(m)}), \end{aligned} \quad (86)$$

and  $\sigma(\cdot)$  denotes a permutation on the set  $[m]$ . Using the chain rule and since  $U_{Y_i} = X_i + N_i$ , we can write  $R_{\sigma}$  as

$$\begin{aligned} &I(X_{\sigma(1)}; U_{Y_{\sigma(1)}} | \mathbf{Y}_{\sigma(1)}) + I(X_{\sigma(2)}; U_{Y_{\sigma(2)}} | \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots \\ &\quad + I(X_{\sigma(m)}; U_{Y_{\sigma(m)}} | \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(m)}), \end{aligned} \quad (87)$$

where each mutual information term is equal to either  $\frac{1}{2} \log \frac{1}{D+\epsilon}$  or 0.

Now we show that  $\max_{\sigma} R_{\sigma}$ , where  $R_{\sigma}$  is equal to (87), equals  $R_{lb}^m(D + \epsilon 1)$ . Note that  $\bar{R}_{lb}$  in Theorem 5 can be written as

$$\max_{\sigma} \bar{R}_{\sigma},$$

where  $\bar{R}_{\sigma}$  is equal to the right-hand side of (26) for a given  $\sigma$ . Using a series of chain rules and expanding the mutual information terms, we can rewrite  $\bar{R}_{\sigma}$  as <sup>15</sup>

$$I(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(1)}; V, U_{Y_{\sigma(1)}} | \mathbf{Y}_{\sigma(1)}) + I(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)}, \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)}; U_{Y_{\sigma(2)}} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots$$

$$+ I(\mathbf{X} \setminus \cup_{i=1}^m \mathbf{Y}_{\sigma(i)}, \cup_{i=1}^m \mathbf{Y}_{\sigma(i)}; U_{Y_{\sigma(m)}} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(m-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(m)}) \quad (88)$$

$$= I(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)}; V, U_{Y_{\sigma(1)}} | \mathbf{Y}_{\sigma(1)}) + I(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)}; U_{Y_{\sigma(2)}} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots$$

$$+ I(\mathbf{X} \setminus \cup_{i=1}^m \mathbf{Y}_{\sigma(i)}; U_{Y_{\sigma(m)}} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(m-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(m)}) \quad (89)$$

$$\stackrel{a}{=} I(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)}; V, U_{Y_{\sigma(1)}} | \mathbf{Y}_{\sigma(1)}) + I(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)}; U_{Y_{\sigma(2)}} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots$$

$$+ I(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)}; U_{Y_{\sigma(k)}} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}) \quad (90)$$

$$\stackrel{b}{=} h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | \mathbf{Y}_{\sigma(1)}) - h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)})$$

$$+ h(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) - h(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, U_{Y_{\sigma(2)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots$$

$$+ h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)})$$

$$- h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}) \quad (91)$$

$$\stackrel{c}{=} h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | \mathbf{Y}_{\sigma(1)}) - h(\mathbf{Y}_{\sigma(2)} \setminus \mathbf{Y}_{\sigma(1)} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}) - h(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)})$$

$$+ h(\mathbf{X} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)})$$

$$- h(\mathbf{Y}_{\sigma(3)} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, U_{Y_{\sigma(2)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) - h(\mathbf{X} \setminus \cup_{i=1}^3 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, U_{Y_{\sigma(2)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}, \mathbf{Y}_{\sigma(3)}) +$$

$$\dots + h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)})$$

$$- h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}) \quad (92)$$

$$= h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | \mathbf{Y}_{\sigma(1)}) - h(\mathbf{Y}_{\sigma(2)} \setminus \mathbf{Y}_{\sigma(1)} | V, U_{Y_{\sigma(1)}}, \mathbf{Y}_{\sigma(1)}) - h(\mathbf{Y}_{\sigma(3)} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, U_{Y_{\sigma(2)}}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) - \dots$$

$$- h(\mathbf{Y}_{\sigma(k)} \setminus \cup_{i=1}^{k-1} \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k-1)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k-1)})$$

$$- h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | V, U_{Y_{\sigma(1)}}, \dots, U_{Y_{\sigma(k)}}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}) \quad (93)$$

$$\geq h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | \mathbf{Y}_{\sigma(1)}) - h(\mathbf{Y}_{\sigma(2)} \setminus \mathbf{Y}_{\sigma(1)} | \hat{X}_{\sigma(1)}, \mathbf{Y}_{\sigma(1)}) - h(\mathbf{Y}_{\sigma(3)} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \hat{X}_{\sigma(2)}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) + \dots$$

$$- h(\mathbf{Y}_{\sigma(k)} \setminus \cup_{i=1}^{k-1} \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(k-1)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k-1)})$$

$$- h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(k)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}), \quad (94)$$

$$= h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | \mathbf{Y}_{\sigma(1)}) - \sum_{j=1}^{k-1} h(\mathbf{Y}_{\sigma(j+1)} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)})$$

$$- h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(k)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}), \quad (95)$$

where  $\hat{X}_{\sigma(i)}$  is such that  $K_{X_{\sigma(i)} | \hat{X}_{\sigma(i)}, \mathbf{Y}_{\sigma(i)}} \leq D + \epsilon$ ,  $N_{\sigma(i)}$  is as defined before, and

a : follows since  $\mathbf{Y}_i = (X_{i-1}, X_{i+1})$  and there exists a  $k > 1$  such that  $\mathbf{X} \setminus \cup_{i=1}^l \mathbf{Y}_{\sigma(i)} = \emptyset$  for all  $l > k$ .

b : follows by expanding each mutual information term.

c : follows by applying the chain rule to all minus terms except the last one (i.e., the second term, the fourth term, etc.).

Now

$$h(\mathbf{Y}_{\sigma(j+1)} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)})$$

$$\leq h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)})$$

$$+ h(\mathbf{X}_{\sigma(j+1)+1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}).$$

$$(96)$$

We shall show that

$$h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)})$$

$$\leq h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | X_{\sigma(1)} + N_{\sigma(1)}, \dots, X_{\sigma(j)} + N_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}) \quad (97)$$

for all  $j \in \{1, \dots, k-1\}$  and similarly for the quantities

$$h(\mathbf{X}_{\sigma(j+1)+1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}). \quad (98)$$

<sup>15</sup>We interpret the differential entropy of an empty set of continuous random variables to be zero.

and

$$h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(k)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}), \quad (99)$$

appearing in (96) and (95), respectively. To show (97), fix  $j \in \{1, \dots, k-1\}$  and define the sets of indices <sup>16</sup>

$$I_1 = \cup_{i=1}^j \{\sigma(i) - 1 \mod m, \sigma(i) + 1 \mod m\} \quad (100)$$

$$I_2 = \cup_{i=1}^j \{\sigma(i)\}. \quad (101)$$

If  $\sigma(j+1) - 1 \mod m \in I_1$ , then the entropy quantities on both sides of (97) are empty so (97) trivially holds. If  $\sigma(j+1) - 1 \mod m \in I_2 \setminus I_1$ , then we have

$$h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}) \quad (102)$$

$$\leq \frac{1}{2} \log(2\pi e(D + \epsilon)) \quad (103)$$

$$= h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | X_{\sigma(1)} + N_{\sigma(1)}, \dots, X_{\sigma(j)} + N_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}). \quad (104)$$

And if  $\sigma(j+1) - 1 \mod m \notin I_1 \cup I_2$ , then we have

$$h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | \hat{X}_{\sigma(1)}, \dots, \hat{X}_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}) \quad (105)$$

$$\leq h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)}) \quad (106)$$

$$= h(\mathbf{X}_{\sigma(j+1)-1} \setminus \cup_{i=1}^j \mathbf{Y}_{\sigma(i)} | X_{\sigma(1)} + N_{\sigma(1)}, \dots, X_{\sigma(j)} + N_{\sigma(j)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(j)}). \quad (107)$$

This establishes (97). The argument for the quantities in (98) and (99) is similar. Substituting into (95) gives

$$\begin{aligned} \bar{R}_\sigma &\geq h(\mathbf{X} \setminus \mathbf{Y}_{\sigma(1)} | \mathbf{Y}_{\sigma(1)}) - h(\mathbf{Y}_{\sigma(2)} \setminus \mathbf{Y}_{\sigma(1)} | X_{\sigma(1)} + N_{\sigma(1)}, \mathbf{Y}_{\sigma(1)}) \\ &\quad - h(\mathbf{Y}_{\sigma(3)} \setminus \cup_{i=1}^2 \mathbf{Y}_{\sigma(i)} | X_{\sigma(1)} + N_{\sigma(1)}, X_{\sigma(2)} + N_{\sigma(2)}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(2)}) - \dots \\ &\quad - h(\mathbf{Y}_{\sigma(k)} \setminus \cup_{i=1}^{k-1} \mathbf{Y}_{\sigma(i)} | X_{\sigma(1)} + N_{\sigma(1)}, \dots, X_{\sigma(k-1)} + N_{\sigma(k-1)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k-1)}) \\ &\quad - h(\mathbf{X} \setminus \cup_{i=1}^k \mathbf{Y}_{\sigma(i)} | X_{\sigma(1)} + N_{\sigma(1)}, \dots, X_{\sigma(k)} + N_{\sigma(k)}, \mathbf{Y}_{\sigma(1)}, \mathbf{Y}_{\sigma(1)}, \dots, \mathbf{Y}_{\sigma(k)}). \end{aligned} \quad (108)$$

Note that this last inequality is an equality when  $V = \emptyset$  and  $U_{Y_{\sigma(i)}} = \hat{X}_{\sigma(i)} = X_{\sigma(i)} + N_{\sigma(i)}$ , implying that  $\bar{R}_\sigma = R_\sigma$ . Hence,

$$R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) = \max_{\sigma} R_\sigma.$$

From (87), we know that  $\max_{\sigma} R_\sigma$  is equal to  $c \frac{1}{2} \log \frac{1}{D+\epsilon}$ , where  $c$  is an integer. Now we find  $c$ . When the permutation  $\sigma(i) = 2i - 1$ , for  $i \in [\frac{m-1}{2}]$  and  $\sigma(i) = 2(i - \frac{m-1}{2})$  for  $i \in \{\frac{m-1}{2}, \dots, m\}$ , we get  $R_\sigma = \frac{m-1}{2} \frac{1}{2} \log \frac{1}{D+\epsilon}$ , implying  $c \geq \frac{m-1}{2}$ . Also, from Theorem 12, we know that  $c \leq \frac{m}{2}$ . Hence  $c = \frac{m-1}{2}$ . Then we have

$$R_{lb}^m(\mathbf{D} + \epsilon \mathbf{1}) = \frac{m-1}{2} \frac{1}{2} \log \frac{1}{D+\epsilon}. \quad (109)$$

Taking  $\epsilon \rightarrow 0$  on both sides gives the result. ■

## APPENDIX C

We first give the definition of  $\epsilon$ -letter typical sequences [17] and then reference results that are useful to prove Theorem 1.

**Definition 6.** Let  $\epsilon > 0$  be given.  $x^n \in \mathcal{X}^n$  is called an  $\epsilon$ -letter typical sequence with respect to  $p_X$  if

$$\left| \frac{1}{n} N(a|x^n) - p_X(a) \right| \leq p_X(a)\epsilon, \text{ for all } a \in \mathcal{X},$$

where  $N(a|x^n)$  denotes the number of times the symbol  $a$  occurs in  $x^n$ . Also  $\mathcal{T}_\epsilon^{(n)}(p_X)$  denotes the set of all  $\epsilon$ -letter typical sequences with respect to  $p_X$ .

**Definition 7.** Let  $\epsilon > 0$  be given.  $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$  is called a jointly typical sequence with respect to  $p_{XY}$  if

$$\left| \frac{1}{n} N(a, b|x^n, y^n) - p_{XY}(a, b) \right| \leq p_{XY}(a, b)\epsilon, \text{ for all } (a, b) \in \mathcal{X} \times \mathcal{Y}.$$

Also  $\mathcal{T}_\epsilon^{(n)}(p_{XY})$  denotes the set of all jointly typical sequences with respect to  $p_{XY}$ .

<sup>16</sup>Recall that here  $x \mod m$  is defined to be in  $[m]$ .

**Definition 8.** Let  $\epsilon > 0$  be given. The set of conditionally typical sequences,  $\mathcal{T}_\epsilon^{(n)}(p_{XY}|x^n)$ , is defined as

$$\mathcal{T}_\epsilon^{(n)}(p_{XY}|x^n) = \{y^n | (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(p_{XY})\}.$$

**Lemma 7.** [17, Theorem 1.1] Let  $0 < \epsilon \leq \mu_X$  where  $\mu_X = \min_{x \in \text{support}(p_X)} p(x)$  and  $X^n \in \mathcal{X}^n$  is drawn i.i.d. with respect to  $p_X$ . Then

$$1 - \delta_\epsilon(n) \leq \Pr[X^n \in \mathcal{T}_\epsilon^{(n)}(p_X)] \leq 1,$$

where  $\delta_\epsilon(n) = 2|\mathcal{X}|e^{-n\epsilon^2\mu_X}$ .

**Lemma 8.** [17, Theorem 1.3] Let  $0 < \epsilon_1 < \epsilon_2 \leq \mu_{XY}$  where  $\mu_{XY} = \min_{(x,y) \in \text{support}(p_{XY})} p(x,y)$  and  $Y^n \in \mathcal{Y}^n$  drawn i.i.d. with respect to  $p_Y$ . If  $x^n \in \mathcal{T}_{\epsilon_1}^{(n)}(p_X)$  then

$$(1 - \delta_{\epsilon_1, \epsilon_2}(n)) 2^{-n(I(X;Y)+2\epsilon_2 H(Y))} \leq \Pr[Y^n \in \mathcal{T}_{\epsilon_2}^{(n)}(p_{XY} | x^n)] \leq 2^{-n(I(X;Y)-2\epsilon_2 H(Y))},$$

where  $\delta_{\epsilon_1, \epsilon_2}(n) = 2|\mathcal{X}||\mathcal{Y}| \cdot e^{-n\frac{(\epsilon_2-\epsilon_1)^2}{1+\epsilon_1}\mu_{XY}}$ .

**Corollary 1.** Let  $0 < \epsilon_1 < \epsilon_2 \leq \mu_{XYZ}$  where  $\mu_{XYZ} = \min_{(x,y,z) \in \text{support}(p_{XYZ})} p(x,y,z)$ .  $Y^n \in \mathcal{Y}^n$  is drawn i.i.d. with respect to  $p_Y$  and  $Z^n \in \mathcal{Z}^n$  is drawn i.i.d. with respect to  $p_Z$ . If  $x^n \in \mathcal{T}_{\epsilon_1}^{(n)}(p_X)$  then

$$\Pr[(Y^n, Z^n) \in \mathcal{T}_{\epsilon_2}^{(n)}(p_{XYZ} | x^n)] \leq 2^{-n((H(Y)+H(Z)-H(Y,Z|X))-2\epsilon_2(H(Y)+H(Z)))}.$$

*Proof:*

$$\begin{aligned} \Pr[(Y^n, Z^n) \in \mathcal{T}_{\epsilon_2}^{(n)}(p_{XYZ} | x^n)] &= \sum_{(y^n, z^n) \in \mathcal{T}_{\epsilon_2}^{(n)}(p_{XYZ} | x^n)} p_Y^n(y^n) p_Z^n(z^n) \\ &\leq 2^{-n(1-\epsilon_2)H(Y)} 2^{-n(1-\epsilon_2)H(Z)} |\mathcal{T}_{\epsilon_2}^{(n)}(p_{XYZ} | x^n)|, \text{ by [17, Theorem 1.1]} \\ &\leq 2^{-n(1-\epsilon_2)H(Y)} 2^{-n(1-\epsilon_2)H(Z)} 2^{nH(Y,Z|X)(1+\epsilon_2)}, \text{ by [17, Theorem 1.2]} \\ &\leq 2^{-n((H(Y)+H(Z)-H(Y,Z|X))-2\epsilon_2(H(Y)+H(Z)))}. \end{aligned}$$

■

**Lemma 9.** [17, Markov Lemma] Let  $0 < \epsilon_1 < \epsilon_2 \leq \mu_{XYZ}$  where  $\mu_{XYZ} = \min_{(x,y,z) \in \text{support}(p_{XYZ})} p(x,y,z)$  and  $(X^n, Y^n, Z^n)$  is drawn i.i.d. with respect to  $p_{XYZ}$  such that  $X \leftrightarrow Y \leftrightarrow Z$ . If  $(x^n, y^n) \in \mathcal{T}_{\epsilon_1}^{(n)}(p_{XY})$  then

$$\begin{aligned} \Pr[Z^n \in \mathcal{T}_{\epsilon_2}^{(n)}(p_{XYZ} | x^n, y^n) | Y^n = y^n] &= \Pr[Z^n \in \mathcal{T}_{\epsilon_2}^{(n)}(p_{XYZ} | x^n, y^n) | Y^n = y^n, X^n = x^n] \\ &\geq 1 - \delta_{\epsilon_1, \epsilon_2}(n) \end{aligned}$$

where  $\delta_{\epsilon_1, \epsilon_2}(n) = 2|\mathcal{X}||\mathcal{Y}||\mathcal{Z}| \cdot e^{-n\frac{(\epsilon_2-\epsilon_1)^2}{1+\epsilon_1}\mu_{XYZ}}$ .

**Lemma 10.** Let  $A$ ,  $B$  and  $C$  denote the events

$$\begin{aligned} &\{\exists \tilde{\mathbf{k}}_{\mathcal{D}_l'} \text{ such that } \tilde{\mathbf{k}}_{\mathcal{D}_l'} \neq \mathbf{k}_{\mathcal{D}_l'}, \tilde{k}_{\mathcal{D}_l'} = k_{\mathcal{D}_l'}, U_{\mathcal{D}_l'}(\tilde{\mathbf{k}}_{\mathcal{D}_l'}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n)\} \text{ and} \\ &\{\exists \tilde{\mathbf{k}}_{\mathcal{D}_l'} \text{ such that } \tilde{k}_{\mathcal{D}_l'} = \mathbf{1}, U_{\mathcal{D}_l'}(\tilde{\mathbf{k}}_{\mathcal{D}_l'}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n)\} \\ &\{(U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n) \in \mathcal{T}_{\epsilon_{l^*+1}}^{(n)}(p)\} \end{aligned}$$

respectively. Then

$$\Pr(A|C) \leq \Pr(B|C).$$

*Proof:* The proof follows the steps in [1, Lemma 11.1]. We start with showing that for a particular set of bin indices  $b_{\mathcal{D}_l'}$ ,

$$\Pr(A|C, k_{\mathcal{D}_l'} = b_{\mathcal{D}_l'} \text{ is chosen at the encoder}) \leq \Pr(B|C, k_{\mathcal{D}_l'} = b_{\mathcal{D}_l'} \text{ is chosen at the encoder}). \quad (110)$$

We can write

$$\begin{aligned}
& \Pr \left( A | C, k_{\mathcal{D}_l'} = b_{\mathcal{D}_l'} \text{ is chosen at the encoder} \right) \\
&= \sum_{b'_{\mathcal{D}_l'}} p(b'_{\mathcal{D}_l'} | b_{\mathcal{D}_l'}) \Pr \left( \exists \tilde{\mathbf{k}}_{\mathcal{D}_l'} \text{ such that } \tilde{k}_{\mathcal{D}_l'} = b_{\mathcal{D}_l'}, \tilde{k}'_{\mathcal{D}_l'} \neq b'_{\mathcal{D}_l'}, U_{\mathcal{D}_l'}(\tilde{\mathbf{k}}_{\mathcal{D}_l'}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n) \right) \\
&\quad \left( C, \mathbf{k}_{\mathcal{D}_l'} = (b_{\mathcal{D}_l'}, \bar{b}'_{\mathcal{D}_l'}) \text{ is chosen at the encoder} \right) \\
&\stackrel{a}{=} \sum_{b'_{\mathcal{D}_l'}} p(b'_{\mathcal{D}_l'} | b_{\mathcal{D}_l'}) \Pr \left( \exists \tilde{\mathbf{k}}_{\mathcal{D}_l'} \text{ such that } \tilde{k}_{\mathcal{S}_j} = 1, \tilde{k}'_{\mathcal{S}_j} \in [2^{R_{\mathcal{S}_j}} - 1] \text{ for all } \mathcal{S}_j \in \mathcal{D}_l', \right. \\
&\quad \left. U_{\mathcal{D}_l'}(\tilde{\mathbf{k}}_{\mathcal{D}_l'}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n) \right) \left( C, \mathbf{k}_{\mathcal{D}_l'} = (b_{\mathcal{D}_l'}, \bar{b}'_{\mathcal{D}_l'}) \text{ is chosen at the encoder} \right) \\
&\leq \sum_{b'_{\mathcal{D}_l'}} p(b'_{\mathcal{D}_l'} | b_{\mathcal{D}_l'}) \Pr \left( \exists \tilde{\mathbf{k}}_{\mathcal{D}_l'} \text{ such that } \tilde{k}_{\mathcal{S}_j} = 1 \text{ for all } \mathcal{S}_j \in \mathcal{D}_l', U_{\mathcal{D}_l'}(\tilde{\mathbf{k}}_{\mathcal{D}_l'}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n) \right) \\
&\quad \left( C, \mathbf{k}_{\mathcal{D}_l'} = (b_{\mathcal{D}_l'}, \bar{b}'_{\mathcal{D}_l'}) \text{ is chosen at the encoder} \right) \\
&= \Pr \left( \exists \tilde{\mathbf{k}}_{\mathcal{D}_l'} \text{ such that } \tilde{k}_{\mathcal{S}_j} = 1 \text{ for all } \mathcal{S}_j \in \mathcal{D}_l', U_{\mathcal{D}_l'}(\tilde{\mathbf{k}}_{\mathcal{D}_l'}) \in \mathcal{T}_{\epsilon_{l^*+2}}^{(n)}(p|U_{\mathcal{D}_l \setminus \mathcal{D}_l'}(\mathbf{k}_{\mathcal{D}_l \setminus \mathcal{D}_l'}), Y_l^n) \right) \\
&\quad \left( C, k_{\mathcal{D}_l'} = b_{\mathcal{D}_l'} \text{ is chosen at the encoder} \right) \\
&= \Pr(B | C, k_{\mathcal{D}_l'} = b_{\mathcal{D}_l'} \text{ is chosen at the encoder}), \tag{111}
\end{aligned}$$

where

a : follows because given any set of codeword indices  $\mathbf{b}_{\mathcal{D}_l'} = (b_{\mathcal{D}_l'}, \bar{b}'_{\mathcal{D}_l'})$  and event  $C$ , for each  $\mathcal{S}_j \in \mathcal{D}_l'$ , any collection of  $[2^{R_{\mathcal{S}_j}} - 1]$ , the number of codewords  $u^n(\mathbf{k}_{\mathcal{S}_j})$  whose index  $\mathbf{k}_{\mathcal{S}_j}$  is different from  $\mathbf{b}_{\mathcal{S}_j}$  has the same distribution.

b: Each bin in codebook  $\mathcal{C}^{\mathcal{S}_j}$  has size  $2^{R_{\mathcal{S}_j}}$ .

Multiplying both sides of (111) with  $p(b_{\mathcal{D}_l'})$  and summing over all bin indices  $b_{\mathcal{D}_l'}$  concludes the proof.  $\blacksquare$

## REFERENCES

- [1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge University Press, 2012.
- [2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1–10, Jan 1976.
- [3] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *Information Theory, IEEE Transactions on*, vol. 31, no. 6, pp. 727–734, 1985.
- [4] S. Watanabe, "The rate-distortion function for product of two sources with side-information at decoders," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, July 2011, pp. 2761–2765.
- [5] R. Timo, T. Oechtering, and M. Wigger, "Source coding problems with conditionally less noisy side information," *Information Theory, IEEE Transactions on*, vol. 60, no. 9, pp. 5516–5532, Sept 2014.
- [6] M. Benammar and A. Zaidi, "Rate-distortion function for a Heegard-Berger problem with two sources and degraded reconstruction sets," *CoRR*, vol. abs/1508.06434, 2015. [Online]. Available: <http://arxiv.org/abs/1508.06434>
- [7] S. Unal and A. B. Wagner, "Vector Gaussian rate-distortion with variable side information," in *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*, 2014.
- [8] —, "Vector Gaussian multi-decoder rate-distortion : Trace constraints," in *Information Sciences and Systems (CISS), 2016 50th Annual Conference on*, 2016.
- [9] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Foundations of Computer Science, 2006. FOCS '06. 47th Annual IEEE Symposium on*, 2006, pp. 197–206.
- [10] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," *Information Theory, IEEE Transactions on*, vol. 59, no. 9, pp. 5811–5823, 2013.
- [11] S. Unal and A. B. Wagner, "A rate-distortion approach to index coding," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6359–6378, Nov 2016.
- [12] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 5240–5257, 2011.
- [13] O. L. Mangasarian and T.-H. Shiau, "Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems," *SIAM Journal on Control and Optimization*, vol. 25, no. 3, pp. 583–595, 1987. [Online]. Available: <http://dx.doi.org/10.1137/0325033>
- [14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL, USA: Academic Press, Inc., 1982.
- [15] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Index coding via linear programming," *ArXiv preprint*, vol. 1004.1379, 2011.
- [16] S. Unal and A. B. Wagner, "An LP lower bound for rate distortion with variable side information," in *Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on*, 2016, pp. 535 – 539.
- [17] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.